

Solar Interior-Atmosphere Linkages and Space Weather Applications

by

Eddie Ross



A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Solar and Stellar Physics Group (SASP)

School of Physics and Astronomy

University of Birmingham

Birmingham, B15 2TT

March 2021

Abstract

The solar cycle gives rise to complex structures and dynamics on the Sun. As a result of magnetic disturbances on the Sun, large bursts of energy lead to space weather events that can be potentially harmful to life on Earth. In this thesis, we present a series of projects exploring the themes of Cosmic Ray (CR) space weather applications and understanding the solar interior-atmosphere linkage.

A feasibility study was performed to determine whether the High School Project on Astrophysics and Research with Cosmics (HiSPARC) network was suitable for detecting space weather events. Using simulations and HiSPARC data, we found few events would be detectable in the nominal HiSPARC configuration, aside from the most extreme events, hence we explored an alternative configuration. We introduced a new configuration of HiSPARC station, which minimised CR energy biases and noise. This configuration was demonstrated to significantly improve the capabilities of HiSPARC as a space weather monitor and we provided compelling evidence to suggest that we should be capable of observing Ground Level Enhancements (GLEs). In addition, we showed the sensitivity would be further enhanced by using a network of detectors.

Analysing long-term variations of Galactic Cosmic Rays (GCRs) versus Sun Spot Sunspot Number (SSN) during recent solar cycles, we explored the relationship between solar activity and GCRs. Focussing on the most recent Cycle 24, we showed it behaved in-accordance with previous even-numbered cycles.

Finally, a frequency-domain analysis of over 20 years of high-cadence Birming-

*T most of
the time in
thesis you
do use it as
1 word.*

ham Solar Oscillations Network (BiSON) observations of the Solar Mean Magnetic Field (SMMF) was presented. This provided evidence to suggest the strongest component of the SMMF is connected to Active Regions (ARs), based both on the inferred lifetime and location on the solar disc. We also searched for evidence of a magnetic signature of global Rossby modes (r modes) in the residual spectrum of the SMMF, where we did not conclusively find an r mode signal in the data.

Acknowledgements

Firstly, I would like to sincerely thank my supervisor, Professor Bill Chaplin. Thank you for inspiring me to enjoy solar and stellar physics as a undergraduate, for your advice and support throughout this Ph.D., and also for allowing me the freedom to explore other creative ventures during my studies.

I would also like to thank my secondary supervisor, Professor Cristina Lazzeroni, for expert support during my Ph.D. and offering me opportunities to get involved in outreach and engagement activities.

I would also like to thank Dr. Angela Romano and Dr. Francesco Gonnella for their valuable support. In addition, thank you to Dr. Steve Hale for advice and for allowing me aid on a maintenance visit to Carnarvon. My experience was made much easier with their guidance.

Thanks to all colleagues within and connected to the Solar and Stellar Physics (SASP) group, or Sun, Stars, and Exoplanets (SSE) group, and within the Astrophysics and Space Research (ASR) group. I would particularly like to thank my incredible fellow Ph.D. students: Vedad, Tom, Ben, Mat, Oli, Alex D., Saniya, Matt, Walter, George, Emma, Alex L., Lindsey, and Emily, for some fantastic memories. The games, drinks, meals out, tea breaks, and friendship all made my experience absolutely delightful. Furthermore, thank you to the postdocs: Ted, Martin, Warrick, Matteo, Jake, and Elinore.

This Ph.D. would not have been possible without the work of the staff in the School of Physics and Astronomy and the wider support staff at the University of

Birmingham. Thank you namely to Kat Grover, Lou Barden, Maria Pavlidou, and Holly Prescott; in addition to the staff in the Physics and Astronomy stores and the education support office.

Thank you to my family for supporting me through the Ph.D.: Mum, Dad, Hayley, Chris, and Connor. Thank you to all my friends, new and old, who have been a part of this with me. An enormous thank you to Dr. Hannah Serrage whose enthusiasm, kindness, and wisdom has helped me tremendously through this odyssey.

Finally, thank you to my reviewers and editors, and my thesis committee.

Contents

Abstract	ii
Acknowledgments	iv
List of Figures	xxi
List of Tables	xxiv
List of Abbreviations	xxviii
1 Introduction	1
1.1 Solar Activity	1
1.1.1 Background	1
1.1.2 Features of Solar Activity	6
1.2 Space Weather	13
1.2.1 Background	13
1.2.2 Impacts of Space Weather	15
1.3 Cosmic Rays	19
1.3.1 Background	19
1.3.2 Cosmic Rays in the Atmosphere	22
1.3.3 Cosmic Ray Detectors	24
1.3.4 Cosmic Ray Observations of Solar Activity and Space Weather	27
1.4 The HiSPARC Experiment	35
1.4.1 Background	35
1.4.2 HiSPARC Detector and Station Configuration	36
1.4.3 HiSPARC Data Acquisition	39
1.5 Thesis Structure	44
2 HiSPARC as a Space Weather Detector	47
2.1 Introduction	47
2.2 Aims	52
2.3 HiSPARC Data	54
2.4 Station Properties	56
2.4.1 Cut-Off Rigidity	56
2.4.2 Asymptotic Viewing Directions	60
2.5 HiSPARC Observations	62
2.5.1 Observations of Ground Level Enhancements	62

2.5.2	Observations of Forbush Decreases	65
2.6	Atmospheric Corrections of HiSPARC Data	69
2.6.1	Motivation	69
2.6.2	Barometric Correction	69
2.6.3	Temperature Correction	72
2.7	HiSPARC Observations After Atmospheric Corrections	75
2.7.1	Observations of Ground Level Enhancements	75
2.7.2	Observations of Forbush Decreases	77
2.8	Air Shower Simulations	79
2.8.1	Motivation	79
2.8.2	Air Shower Footprints	80
2.8.3	Muon Flux	82
2.8.4	Muon Flux From MAIRE	84
2.9	Conclusion	92
3	HiSPARC Station 14008	96
3.1	Introduction	96
3.2	Aims	97
3.3	HiSPARC Station 14008 Set-Up	97
3.3.1	Configuration	97
3.3.2	Calibration	102
3.3.3	Monitoring Temperature	104
3.3.4	Data Acquisition	104
3.3.5	Monitoring Pressure	107
3.4	Methodology	108
3.4.1	Atmospheric Corrections	108
3.4.2	Observation Statistics	109
3.5	Atmospheric Corrections	113
3.5.1	Temperature Correction	113
3.5.2	Barometric Correction	116
3.6	Results	118
3.6.1	Observations	118
3.6.2	Comparison with Neutron Monitors	123
3.6.3	Single Station Space Weather Uses	125
3.6.4	Multiple Station Space Weather Uses	131
3.7	Conclusions	137
4	Galactic Cosmic Ray Behaviour During Solar Cycle 24	141
4.1	Introduction	141
4.2	Data	146
4.3	Time-Lag Analysis	148
4.3.1	Method	148
4.3.2	Results	148
4.4	Hysteresis Effect Analysis	153
4.4.1	Method	153
4.4.2	Results	154
4.5	Conclusions	159

4.6 Comparison with HiSPARC	160
5 A Frequency Domain Investigation on the Morphology of the Solar Mean Magnetic Field	163
5.1 Introduction	163
5.2 Aims	167
5.3 Data	168
5.3.1 Summary of the Data Set	168
5.3.2 Obtaining the SMMF from BiSON	169
5.4 Methodology	172
5.4.1 Identifying Features in the SMMF Power Spectrum	172
5.4.2 Parametrisation of the SMMF Power Spectrum	175
5.4.3 Modelling the SMMF Power Spectrum	177
5.4.4 Comparison with the WSO SMMF	178
5.5 Results	179
5.5.1 Investigation of the Window Function	179
5.5.2 Modelling the BiSON Power Spectrum	184
5.5.3 Comparison to the WSO Power Spectrum	188
5.6 Discussion	192
5.6.1 Testing the Effects of Differential Rotation and Active Region Migration	192
5.6.2 Further Morphology of the SMMF using SDO/HMI Data	196
5.7 Conclusion	199
6 Rossby Modes in the Solar Mean Magnetic Field	203
6.1 Introduction	203
6.2 Theory	205
6.3 Methodology	207
6.3.1 Testing the Residual Spectrum	207
6.3.2 Modelling r mode Profiles	210
6.4 Results	211
6.4.1 Testing the Residual Spectrum	211
6.4.2 Modelling r mode Profiles	213
6.5 Discussion	215
6.5.1 Manifestation of Rossby Waves in the Power Spectrum	215
6.5.2 Rossby Modes in Other Sources of SMMF Data	220
6.6 Conclusion	221
7 Conclusions and Future Prospects	224
A Simulations of Cosmic Ray Air Showers using CORSIKA	232
B Simulations of the Ground Level Enhancements for the Analysis of HiSPARC Station 14008	236
B.1 Model	236
B.2 Configuration of the Simulations	238

List of Figures

1.1	Schematic diagram of the main processes that drive the changes to the Sun’s magnetic field. Starting from a poloidal field on the left, it is shown how the Ω -effect generates a toroidal field. On the top we see a demonstration of the creation of small-scale poloidal magnetic fields, the α -effect. On the bottom we see a demonstration of the Babcock-Leighton mechanism, diffusion and meridional flows. Finally, on the right we see the recovery of a poloidal field with opposite polarity. Adapted from Sanchez et al. (2014)	2
1.2	Monthly averaged sunspot number since 1750 (Hathaway & Upton, 2017). Black points indicate a full month of data, and then a traffic light system is in place to represent data quality, i.e. green for a near-complete month of data, amber for near 50% duty cycle, and red for a very low data duty cycle.	4
1.3	Top panel: shows the latitudinal distribution of sunspots over time (also known as a “Butterfly Diagram”) and the colour of the points depict the area of the disc. Bottom panel: shows the average daily area of sunspots on the visible solar disc. Taken from Hathaway & Upton (2017)	5
1.4	UK National Risk Register for hazards, diseases, accidents, and societal risks showing space weather as a medium-high risk (Cabinet Office, 2017)	16
1.5	The sources and effects of space weather. Impacts are shown including loss of telecommunications and GNSS, increased radiation levels, and ground induced currents (ESA/Science Office, CC BY-SA 3.0 IGO) . .	16
1.6	Cosmic ray differential energy spectrum using data measured by several experiments. The inset shows the H/He ratio at constant rigidity (Particle Data Group et al., 2020)	21

1.7	Schematic diagrams of two NM configurations: (a) the original Simpson's 12-tube IGY NM is shown where the paraffin reflector is represented by the outer dotted blocks; the lead producer is represented by the cross-hatched section; the paraffin moderator is represented by the inner dotted blocks; finally the gas-filled proportional counters are denoted by the black circles/tubes. (b) the modern NM64 is shown on the right in its 6-tube configuration where the polyethylene reflector is represented by the outer dotted section; the lead producer is a cylindrical polyethylene moderator denoted by the dotted ring; finally in the centre of each tube is a gas-filled proportional counter. In each figure the top schematic is a top-down drawing and the bottom is an end-on drawing. Taken from Kang et al. (2012).	25
1.8	Schematic design of a typical scintillation muon detector with back-end electronics (Gloeckler, 2010)	26
1.9	Cosmic ray flux measured at the Climax NM (solid line) and rescaled SSN (dotted line) taken from Hathaway (2015). The vertical lines denote changes in the Sun's global magnetic field polarity, where: A+ indicates positive polarity at the North Pole; A- indicates negative polarity at the North Pole.	28
1.10	A two-step Forbush decrease measured at three NM stations, Deep River, Mt. Wellington, Kerguelen, in July 1982 (Cane, 2000). The thicker black line indicates the average of the count rates from the three stations. Arrows show the start of the two decreases caused by the shock and the ICME ejecta.	29
1.11	A schematic diagram of an ICME-driven FD taken from Cane (2000). It shows the different cosmic ray responses from two paths, indicated by A and B. Path A experiences the shock and ejecta, therefore experiencing a two-step FD; B only experiences the shock, therefore experiencing a single decrease. The time of shock passage is indicated by a solid, vertical line marked, S; the start and end times of ejecta passage are indicated by vertical, dashed lines marked T1 and T2, respectively.	30
1.12	A GLE measured at nine NM stations in October 1989, taken from Cramp, J. L. et al. (1997).	31
1.13	A schematic diagram of the 'garden hose' field line taken from Duldig et al. (1993).	33
1.14	Schematic diagram of the HiSPARC scintillation detector. (A): PMT; (B): light-guide adaptor; (C): light-guide; (D): scintillator.	36
1.15	Typical formations of two-detector and four-detector stations (Fokkema, 2012; van Dam et al., 2020b). In each, the black circle denotes a GPS antenna which is located in between the detectors to provide a precise timestamp for each signal.	38
1.16	A schematic diagram showing the configuration and interfaces between the HiSPARC hardware for a two-detector station.	40

1.17 (a): An example PMT signal after digital conversion by the HiSPARC electronics box. The horizontal lines denote: the noise cut-off (dotted line), which is used for setting a limit when integrating the pulse height, to give the pulse integral; the low-voltage threshold (dash-dot); the high-voltage threshold (dashed). The role of the high- and low-voltage thresholds are described in the text below. (b) The pulse height distribution over the course of a single day from HiSPARC station 501. The vertical lines show the low-voltage threshold (dash-dot) and the high-voltage threshold (dashed).	40
1.18 Schematic data acquisition of an event (Fokkema, 2012). The dashed vertical lines denote the epochs of the pre-trigger, coincidence, and post-trigger windows. The grey, shaded region shows the data reduction window and data outside this window are not stored. The dotted, horizontal lines denote the low- and high-voltage thresholds.	43
2.1 GLEs observed by the NM stations based at Oulu. Top panel: GLE 70; middle panel: GLE 71, bottom panel: GLE 72. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-red lines show the epochs of each GLE onset. The units of time on the x-axis are, MM-DD HH.	52
2.2 FDs observed by the NM stations based at Oulu. Top left panel: FDs during March 2012; top right panel: FD during July 2012; bottom left panel: FD during December 2014; bottom right panel: FD during September 2017. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-blue lines show the epochs of each FD onset. The units of time on the x-axis are, YYYY-MM-DD.	53
2.3 The geographic location of each HiSPARC station considered in this work. Each green circle denotes the location of a detector station.	56
2.4 The availability of data for each HiSPARC station considered, for each of the space weather epochs listed in Table 2.1. The purple grids denote no available data, teal denotes that only the events data are available, and yellow denotes that both the events and singles data are available.	57
2.5 Azimuthal and zenith angle variations in the allowed and forbidden rigidity trajectories for HiSPARC station 501 from simulations in steps of rigidity, $\Delta R = 0.01$ GV. The forbidden trajectories are in black; allowed trajectories are in white.	60
2.6 The vertical asymptotic viewing directions of 5 HiSPARC stations. The rigidity range of the simulations were from $1.0 \text{ GV} < R < 20.0 \text{ GV}$, and the results are plotted in geographic coordinates on January 20th 2005. The diamonds correspond to the HS ground location and the circles correspond to the AVD for a specific rigidity value.	61

2.7	HiSPARC data for stations 501 and 3001 around the epoch of GLE 70 on 13/12/2006. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.	62
2.8	HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.	62
2.9	HiSPARC data for 4 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the 1-minute (grey) and 5-minute (black) averaged trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH _n) in the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.	63
2.10	HiSPARC data for stations 501 and 8001 around the epoch of the FDs in March 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed lines show the approximate onset-time of the FDs. The units of time on the x-axis are, YYYY-MM-DD.	65
2.11	HiSPARC data for stations 501 and 8001 around the epoch of the FD in July 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.	66
2.12	HiSPARC data for four stations around the epoch of the FD in December 2014. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.	66
2.13	HiSPARC data for four stations around the epoch in which there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station, while the bottom panel shows the minute- and hourly-averaged singles counts, mean-subtracted, for each individual detector (or channel, CH _n) in the station. The vertical blue-dashed lines show the approximate onset-times of the two FDs observed around this epoch and the red-dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, YYYY-MM-DD.	68

2.14 The anti-correlation between CR count rates and the atmospheric pressure. (a) shows the CR and the local atmospheric pressure measured at a NM in the South Pole; (b) shows the CR and pressure measured by HiSPARC station 501. In both plots, the data shown are hourly-averaged, to highlight the effects. The units of time on the x-axis are, YYYY-MM-DD.	70
2.15 The barometric coefficient calculation: (a) during November 2017 for the South Pole (SOPO) NM station, (b) during November 2019 for HiSPARC station 501 at Nikhef.	71
2.16 A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the pressure correction (blue line). The units of time on the x-axis are, YYYY-MM-DD.	72
2.17 A comparison between the monthly barometric coefficient computed in this work and using the online barometric coefficient tool throughout the year 2017 for the SOPO NM station. The units of time on the x-axis are, YYYY-MM.	72
2.18 The relationship between the pressure corrected events data and the outdoor temperature as measured at HiSPARC station 8001 (Eindhoven). (a) shows the time-series of hourly-averaged pressure corrected events and temperature data; (b) shows the correlation between the counts and temperature, and the fitted line to calculate the correction coefficient.	74
2.19 A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the temperature correction (blue line). After the correction, the diurnal variation from the temperature effect has been reduced.	75
2.20 Atmospheric-corrected HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.	76
2.21 Atmospheric-corrected HiSPARC data for 2 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH _n) in the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.	76
2.22 Atmospheric-corrected HiSPARC data for station 8001 (Eindhoven) around the epoch of the FD in July 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.	77

2.23 Atmospheric-corrected HiSPARC data for station 501 (Nikhef) around the epoch of the FD in December 2014. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.	78
2.24 Atmospheric-corrected HiSPARC data for 2 stations in an epoch where there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute-averaged (grey) and hourly-averaged (black) trigger events. The bottom panel shows the minute- and hourly-averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH _n) in the station. The vertical blue-dashed lines show the approximate onset-times of the FDs and the red-dashed line depicts the approximate onset-time of the GLE. The units of time on the x-axis are, YYYY-MM-DD.	78
2.25 Mean muon density footprints for (a) proton-initiated air showers and (b) α -particle-initiated air showers with initial PCR trajectories with zenith angles $\theta = 0^\circ$ and various PCR energies. The error bars given represent 1σ	80
2.26 Mean number of muons produced at ground level from (a) proton-initiated air showers and (b) α -particle-initiated air showers, with initial PCR trajectories with zenith angles $\theta = 0^\circ$ and various PCR energies.	82
2.27 Ground level muon spectra as computed in the CORSIKA simulations. (a) shows the differential muon flux at ground level; (b) shows the muon flux through a single HiSPARC detector. In both plots the solid, blue line shows the simulations using vertically incident PCRs, and the dashed, orange line shows the simulations using PCRs incident within a cone of 70°	83
2.28 Plots of the calculated MAIRE muon spectra for different incident PCR spectra. Blue and orange lines show muon spectra calculated for the incident GCR spectra during solar minimum and maximum, respectively. The other coloured lines show the computed muon spectra for the incident GLE spectra. (a) shows the individual muon spectra for GCRs and GLEs; (b) shows the combined muon spectra for the GCR at solar maximum and the GLEs.	85
2.29 Shows the relationship between the predicted magnitude of GLE 42 and geographical location. Blue square shows the predicted muon increase at the Kiel station, orange up-triangle for HiSPARC station 501 (Nikhef), and green down-triangle for HiSPARC stations 14001 (University of Birmingham). The black points show the results when varying one property of the Kiel simulations, in the top panel we vary latitude; middle panel the variation with longitude; the bottom panel shows the variation with altitude.	89

3.1	Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box.	98
3.2	HiSPARC 14008 assembly and configuration. (a) Shows the stacked arrangement of the scintillators within the roof box, between layers of protective foam. (b) Shows the complete detector on the roof of the Poynting building on the University of Birmingham campus.	99
3.3	Schematic diagram of the HiSPARC 14008 station NIM crate configuration. Black-outlined boxes indicate the NIM crate modules, while red-outlined boxes depict HiSPARC hardware modules. Black arrows depict a NIM signal; green arrows show a TTL signal; blue arrows depict the analogue signal from the PMTs. The PMT signals are split and half of the signal interfaces directly with the HiSPARC electronics box and half the signal is passed through the NIM crate for processing.	100
3.4	Voltage calibration curve for the PMTs of station 14008. The upper, red-dashed line indicates the upper limit for the low threshold singles rate (400 Hz), and the lower 2, black-dashed lines indicate the upper and lower bounds for the high threshold singles rate (100–130 Hz).	103
3.5	(a) Schematic diagram of the DS18B20 temperature sensor circuit, whereby the voltage drain (VDD), ground (GND), and data (DQ) pins connect directly to the voltage supply (+V), ground, and input/output (GPIO) pins of the Raspberry Pi board. (b) Shows the temperature sensor within the roof box, located by the PMTs. The temperature sensor is soldered into the circuit board seen in the top-middle of the image.	105
3.6	Schematic diagram of the HiSPARC station 14008 data acquisition interfaces. Red lines depict a 5 V signal; black lines show ground connection; blue lines depict the 1-wire protocol signal.	105
3.7	The relationship between the original coincidences data and the temperature within the roof box over a single day. (a) Shows the comparison between coincidences data (orange) and temperature data (blue), where the units of time on the x-axis are, MM-DD HH; (b) shows the correlation between the coincidences counts and temperature, and the fitted line to calculate the correction coefficient.	113
3.8	The relationship between the singles data (blue and orange lines) and the temperature within the roof box (black line). The units of time on the x-axis are, YYYY-MM-DD.	114
3.9	The relationship between the singles data and the temperature within the roof box over a single day. (a) Shows the comparison between singles data (orange) and temperature data (blue), where the units of time on the x-axis are: MM-DD HH; (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the correction coefficient.	115

3.10	The singles data before (orange line) and after (blue line) the temperature correction process. The hourly resampled data are over-plotted to highlight the main variation in the data. The units of time on the x-axis are, YYYY-MM-DD.	116
3.11	Comparison between HiSPARC station 14001 singles data and the singles data and temperature measured by station 14008. Black line: the temperature within the roof box of station 14008; dashed, blue line: HiSPARC station 14008 singles data; solid, orange, green, red, and purple lines: HiSPARC station 14001 singles data. The units of time on the x-axis are, YYYY-MM-DD.	117
3.12	The relationship between the smoothed, original coincidences data and the atmospheric pressure. (a) Shows the comparison between coincidences data (orange) and pressure data (blue), both with a 12-hour box-bar smoothing applied, to highlight the relationship and the units of time on the x-axis are: YYYY-MM-DD; (b) shows the correlation between the coincidences counts and pressure and the fitted line to calculate the correction coefficient.	117
3.13	Showing the coincidences data before (orange line) and after (blue line) the barometric correction. The hourly resampled data are over-plotted to highlight the main variation in the data.	118
3.14	Time series of coincidences data, corrected for atmospheric pressure. The blue line shows the corrected data displaying the diurnal variation with peaks at around midday. The orange line shows the data smoothed using a 6-hour box-car. The units of time on the x-axis are, MM-DD HH.	119
3.15	(a) Comparison of the reduced coincidences stored locally (orange) and as events data in the HiSPARC server (blue), where the units of time on the x-axis are, YYYY-MM-DD. (b) Distribution of the ratio of the original and reduced coincidence rates.	121
3.16	Time series of spurious coincidences data measured over five days. The units of time on the x-axis are, YYYY-MM-DD.	122
3.17	Distribution plot of the random coincidences data (hatched bars), and probability mass function of a Poisson distribution with mean equal to the median value of the posterior distribution after sampling (blue bars). Blue error bars represent the 68% credible intervals either side of the median.	123
3.18	A comparison between the corrected HiSPARC station 14008 coincidences data (blue) and the pressure corrected neutron monitor data measured at the Dourbes NM station, Belgium. (a) Shows the two data sets offset and (b) overlayed hourly-averaged data, showing the similarities between the two data sets. Units of time on the x-axis are, YYYY-MM-DD.	124

- 3.19 Single realisation of a simulation with a 10% GLE magnitude. Top panel shows the raw, artificial data from the simulation, with a 10-second cadence. Middle panel shows the number of measurements exceeding the $Z = 3$ and $Z = 5$ thresholds. The $Z = 3$ and $Z = 5$ thresholds are depicted as dotted and dashed lines, respectively. Bottom panel shows the number of points exceeding the $p = 10\%$ threshold in the binomial test, where a low probability value indicates a high chance the measurement is not statistical fluctuation. The dashed line shows the $p = 10\%$ threshold. 126
- 3.20 Summary plots showing the number of excessive measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for the 10-s cadence data; (b) 1-minute averaged data; (c) 5-minute averaged data. In each plot, the top panel shows the number of significant points exceeding the binomial $p = 10\%$ threshold, the middle panel shows the number of points exceeding the $Z=3$ threshold, and the bottom panel shows the number of points exceeding the $Z=5$ threshold. In each panel the dashed, horizontal lines show the median values of the tests without an injected GLE, and the horizontal, dotted lines represent the 68% credible intervals either side of the median. For each simulated GLE magnitude the point represents the median values and the error bars represent the 68% credible intervals either side of the median. 129
- 3.21 Summary plots showing the number of excessive measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for a single station; (b) 2 stations; (c) 5 stations; (d) 10 stations. In each plot, the top panel shows the number of excessive points exceeding the binomial $p = 10\%$ threshold, the middle panel shows the number of points exceeding the $Z=3$ threshold, and the bottom panel shows the number of points exceeding the $Z=5$ threshold. In each panel the dashed, horizontal lines show the median values of the tests without an injected GLE, and the horizontal, dotted lines represent the 68% credible intervals either side of the median. For each simulated GLE magnitude the point represents the median values and the error bars represent the 68% credible intervals either side of the median. 133
- 3.22 Cross-correlation analyses between 2, 5, and 10 stations for a 2% GLE magnitude with varying rise and decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the red line shows the mean of all the realisations. Finally, the dashed, horizontal lines depict a correlation of zero. 135

3.23	Cross-correlation analyses between 2, 5, and 10 stations for a 1% GLE magnitude with varying rise and decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the red line shows the mean of all the realisations. Finally, the dashed, horizontal lines depict a correlation of zero.	136
4.1	SSN (top), with vertical lines showing the beginning of each solar cycle. Cosmic ray intensity recorded by NMs (bottom), with vertical lines showing the approximate epochs of solar magnetic field polarity reversals. (MCMD = McMurdo, NEWK = Newark, SOPO = South Pole, THUL = Thule).	142
4.2	Hysteresis plots between yearly averaged SSN and yearly averaged GCR intensity for each of the 4 main NM stations over cycles 19–24.	145
4.3	Variation in the correlation coefficient with time-lag NM station GCR intensity and SSN during solar cycles 20–23.	149
4.4	Variation in the correlation coefficient with time-lag between NM GCR intensity and SSN during solar cycle 24.	150
4.5	Variation in the correlation coefficient with time-lag between NM GCR intensity and SSN between 2000–2012.	151
4.6	Variation in time-lag plotted against NM station rigidity cut-off for the 16 NM stations detailed in Table 4.1.	153
4.7	The hysteresis plots for even solar cycles 20 and 22 and the linear regression fit to the data.	156
4.8	The hysteresis plots for odd solar cycles 21 and 23 and the linear regression fit to the data.	156
4.9	The hysteresis plots for odd solar cycles 21 and 23 and the ellipse fit to the data.	157
4.10	The hysteresis plot for solar cycle 24, and the linear regression fit to the data (left) and ellipse fit to the data (right).	158
4.11	The monthly-mean CR-induced muon count rate recorded by HiSPARC (blue), and the SSN (black) between 2008 – 2019.	161
4.12	The evolution of the mean-shifted PMT operating voltages for each detector within HiSPARC station 501 between 2010 – 2020.	162
5.1	An example of the BiSON ratios data over a 30-minute period. The separation between the two ratios is due to the solar mean magnetic field. Other excursions in the individual ratios are due to the other effects measured by the RSS.	171
5.2	(a) 40-second cadence observations of the SMMF from the Sutherland BiSON station between 1992 and 2012. The sense of the field was chosen to match the Chaplin et al. (2003) and the WSO observations, where positive is for a field pointing outwards from the Sun. (b) Power spectrum of the SMMF on a 40-second cadence truncated to 10 μ Hz, however, the Nyquist frequency is 12500 μ Hz.	173

5.3	Power spectrum of 40-second cadence SMMF from the Sutherland BiSON station observed between 1992 – 2012 on a logarithmic scale up to the Nyquist frequency.	174
5.4	Locations of aliased power in side-band peaks. The orange, dotted-lines show the locations of frequencies at multiples of 1/day. The green, dashed-lines show the locations of the side-band peaks – harmonic frequencies reflected around multiples of 1/day. The inset shows a zoom of one set of side-band peaks around 1/day.	180
5.5	The effects of the window function on the power spectrum is shown by using a fake data set and this is compared to the BiSON power spectrum. Black line: BiSON SMMF PSD; blue line: power spectrum of the window function; green and dark-orange lines: the power spectrum of the artificial data without and with gaps, respectively; light orange line: the input peak used to generate the artificial data over-plotted. The power spectra of the BiSON SMMF and the window function have been shifted upwards by a factor of 6 and 30, respectively, for clarity.	181
5.6	Full, modelled power spectrum of the BiSON SMMF on logarithmic axes. The data are displayed in black and the convolved model using symmetric Lorentzian peaks is shown in green.	187
5.7	Modelled power spectrum of (a) the WSO SMMF; (b) the daily-averaged BiSON SMMF, on logarithmic axes. The data are displayed in black and the convolved model using asymmetric Lorentzian peaks is shown in blue and green, for WSO and BiSON, respectively.	190
5.8	A comparison between the power spectra produced using the daily-averaged BiSON data and the 40-second cadence BiSON observations. The top plot shows the log-smoothed power spectra of the daily-averaged data (blue) and the 40-second data (orange). The bottom panel show the ratio of the daily-averaged data power spectrum to the 40-second data power spectrum. The horizontal, dashed line indicates a ratio of 1.	191
5.9	(a) Shows the Lorentzian distribution peak before and after the time-averaged broadening, and the fit to the broadened peak. (b) Shows the peak distribution before and after the analytical broadening, and the fit to the broadened peak. In both plots the broadened peaks have been shifted by the relevant frequency to overlay them on top of the true ν_0 for comparison.	195
5.10	Investigations of timescales in the SDO/HMI magnetograms over 2011 and 2014. Both plots show in the top panel, the hemispheric Mean Magnetic Field (MMF) and full-disc SMMF from the magnetograms. The lower panel of each plot displays a comparison between the hemispheric and full-disc mean of the synoptic charts, compared to the box-car smoothed MMF from the magnetograms. N: Northern hemisphere; S: Southern hemisphere. Full HMI: considers the full solar disc; 99 HMI: considers only the inner 99% of the solar disc, by radius.	198

6.1	Full, modelled power spectrum of the BiSON SMMF on logarithmic axes. The data is displayed in black and the convolved model using symmetric Lorentzian peaks is shown in green.	205
6.2	Residual power spectrum of the BiSON SMMF. Over plotted in the dark blue curve is the model of the main SMMF signal which was divided out from the raw spectrum. Also over plotted as vertical solid lines are the expected locations of the 4 lowest-frequency sectoral r modes and the dashed lines, the locations of the B_0 variation frequency splitting. Dashed lines represent ± 31.7 nHz (i.e. representing the frequencies of splitting due to the variation in the B_0 angle.) . . .	208
6.3	Realisations of the statistics tests on the BiSON data for different re-binning factors (n). The panels of each sub figure are: (top) the full PSD and fit, (second panel) the full and re-binned residuals, (third panel) the probability of statistical noise in each bin, (bottom) distribution of the residuals, blue, compared to a χ^2 2n-DOF, orange. . .	212
6.4	Model fit to the r mode in the BiSON PSD residuals using the median values from the posterior distributions of each parameter.	214
6.5	Mode displacement schematic for an $l = m = 2$ r mode (Strohmayer & Mahmoodifar, 2014)	216
6.6	Time series of the velocity and amplitude modulation toy model simulations. The blue curve shows the velocity modulation, i.e. modulating using a cosine with period of 1 year, whereas the orange curve shows the amplitude modulation, i.e. modulating using a rectified cosine with period 1 year.	217
6.7	Power spectra for the two modulation methods, showing the difference in the way the modulation has changed the frequency of the observed mode. The dashed, vertical line shows the location of the central frequency and the dotted, vertical lines show multiples of ν_{\oplus} from the central frequency.	217
6.8	SDO/HMI SMMF split into hemispheres and compared to other SMMF sources during (a) 2011, and (b) 2014. The top panel in each figure shows the north (N), south (S), and total disc-averaged mean magnetic field, for both the full solar disc and from pixels within 99% of solar radius. The bottom panels show a comparison between the SMMF, as observed BiSON, WSO, and SDO/HMI (full disc and the 99% disc).	219
6.9	Comparison of the power spectra for BiSON, WSO, and SDO/HMI. In both figures, the top panel shows the BiSON PSD and the bottom panel shows either the WSO or HMI PSD. The dashed, black line shows the location of the theoretical $l = 2 = m$ r mode frequency, and the red, solid line shows the location of the peak fit in the BiSON PSD residuals.	221

B.1	Example of a realisation of the simulations, to generate an artificial time series with a 10% GLE injected. The black line shows the model for the GLE. The blue line shows the realisation of the signal without a GLE injected, and the orange line shows the realisation of the model with the GLE injected.	237
B.2	Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.	238

List of Tables

1.1	Most prominent decay modes of the mesonic components of CR air showers and of muons. Note: K^- modes are charge conjugates of the decay modes below (Particle Data Group et al., 2020)	23
2.1	Largest space weather events since the beginning of HiSPARC, which were searched for within the HiSPARC data. The percentage-change columns provide a reference of how much the CR counts observed by the NM stations at Oulu ($R_C=0.81$ GV) and Kiel ($R_C=2.36$ GV) increased by or decreased by, due to the space weather event. More precise times for the event onset can be found within: NMDB (2018) (for GLEs) and Lingri et al. (2016) (for FDs).	51
2.2	Properties of some of the HiSPARC stations: geographic longitude (ϕ), geographic latitude (λ), altitude (h), and the geomagnetic vertical cut-off rigidity (R_C) calculated from the PLANETOCOSMICS simulations.	59
2.3	The seven GLEs where MAIRE muon spectra were available, and the maximum observed increase in the neutron flux in the NMDB and the station where the increase was observed.	85
2.4	Observed and predicted increases in the CR count rates at the Kiel NM station, with rows ordered by the observed GLE magnitude. The observed neutron increases use data from the NMDB, where the errors are the measurement uncertainty from the 5-minute averaged data. The predicted data for the neutrons and muons are from the MAIRE simulations. The ratio column provides a conversion factor from the MAIRE predicted neutron increases compared to those observed. The fraction column shows the proportion of the MAIRE predicted muon increase compared to the MAIRE predicted neutron increase.	87
2.5	The MAIRE predicted increase in the muon flux at Kiel and two HiSPARC stations. The rows are ordered by the observed NM magnitude from Table 2.4.	88
2.6	The MAIRE predicted increase in the muon flux, adjusted for the calibration factor between observed and predicted neutron increases at the Kiel station, in Table 2.4. The rows are ordered by the observed NM magnitude from Table 2.4.	91

3.1	Delays in the signals through different paths in the NIM set-up. The paths all start from the output of the PMTs, and are either direct or pass though the NIM crate before reaching their final end interface, therefore the path column is formatted as: start – direct/NIM path – end.	102
3.2	Variables stored in the coincidences files of the HiSPARC 14008 instrument.	106
3.3	Variables stored in the temperature files of the HiSPARC 14008 instrument.	106
3.4	Median number of excessive measurements in the artificial data with no injected GLEs. The values were acquired from 1000 iterations of simulated data, performing each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for the 10-s cadence data and averaging over 1-minute and 5-minutes. All uncertainties correspond to the 68% credible intervals either side of the median.	127
3.5	Median number of excessive measurements in the mean of the artificial data for multiple stations, with no injected GLEs. The values were acquired from 1000 iterations of simulated data, performing each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, with 1, 2, 5, and 10 stations. All uncertainties correspond to the 68% credible intervals either side of the median.	132
4.1	Neutron monitor stations used in this study and their vertical geomagnetic cut-off rigidity (R_c), longitude, latitude, and altitude acquired from NEST. The first four stations have been used for all of the analysis while the lower 12 stations have been used exclusively for the investigation into the dependence of R_c on the time-lag.	147
4.2	Time-lags and the corresponding cross-correlation coefficient between NM CR count and SSN for solar cycles 20–23.	150
4.3	Time-lags and the corresponding cross-correlation coefficient between NM GCR intensity and SSN for solar cycle 24.	150
4.4	Time-lags and the corresponding cross-correlation coefficient between NM GCR intensity and SSN during 2000–2012.	151
4.5	Correlation coefficients of the linear regression and ellipse modelling of the hysteresis plots for solar cycles 20–23.	155
4.6	Correlation coefficients of the linear regression and ellipse modelling of the hysteresis plots for solar cycle 24.	157
5.1	Model parameter values for the generation of artificial data, and the median posterior values for the fit to the power spectra generated with and without the gaps in the data. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.	183

5.2	Median values of the marginalised posterior distributions for each model parameter in the fit to the BiSON power spectrum using symmetric and asymmetric Lorentzian profiles. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median in the adjusted posteriors. The last row in the table shows the Bayesian Information Criterion (BIC) value for each model.	186
5.3	Median values of the marginalised posterior distributions for each model parameter in the fit to the daily WSO and BiSON power spectra. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median. The last row in the table shows the BIC value for each model.	190
5.4	Input linewidth and the median posterior values of the Lorentzian model each simulation. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.	195
6.1	Predicted ⁺ and observed [◦] r mode frequencies for combinations of l and m . Predicted frequencies and conversions of observations to different frames of reference use equation (6.1), equation (6.2), and equation (6.3), with $\Omega = 453.1$ nHz. The predicted splitting for the B_0 angle variation is also provided. The key for the source column is: LPT for Löptien et al. (2018), LNG for Liang et al. (2019), and LZA for Lanza et al. (2019).	207
6.2	Median posterior values of the Lorentzian model for the r mode peak in the BiSON SMMF PSD. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.	213
A.1	Details of the proton-initiated air showers simulated using CORSIKA.	234
A.2	Details of the α -particle-initiated air showers simulated using CORSIKA.	235

List of Abbreviations

AMS-02 Alpha Magnetic Spectrometer.

AR Active Region.

AS Air Shower.

AVD Asymptotic Viewing Direction.

BIC Bayesian Informatin Criterion.

BiSON Birmingham Solar Oscillations Network.

BMR Bipolar Magnetic Region.

CCF Cross-Correlation Function.

CEDA Centre for Environmental Data Analysis.

CIR Corotating Interaction Region.

CMA Central Moving Average.

CMB Cosmic Microwave Background.

CME Coronal Mass Ejection.

CORSIKA Cosmic Ray Simulations for Kascade.

CR Cosmic Ray.

DAQ Data Acquisition.

EAS Extensive Air Shower.

EEMD Empirical Mode Decomposition.

ER Ephemeril Region.

ESD Event Summary Data.

FA False Alarm.

FD Forbush Decrease.

FE Forbush Effect.

FEID Forbush Effects and Interplanetary-disturbances Database.

FPGA Field Programmable Gate Array.

FWHM Full Width at Half Maximum.

GCR Galactic Cosmic Ray.

GIC Ground Induced Current.

GLE Ground Level Enhancement.

GMF General Magnetic Field.

GNMN Global Neutron Monitor Network.

GNSS Global Navigation Satellite System.

GPS Global Positioning System.

GZK Greisen-Zatsepin-Kuzmin.

HCS Heliospheric Current Sheet.

HiSPARC High School Project on Astrophysics and Research with Cosmics.

HMC Hamiltonian Monte Carlo.

HMF Heliospheric Magnetic Field.

HV High Voltage.

ICME Interplanetary Coronal Mass Ejection.

IGRF International Geomagnetic Reference Field.

IMF Interplanetary Magnetic Field.

ISS International Space Station.

ISSN International Sun Spot Number.

LOS Line Of Sight.

MAIRE Model for Atmospheric Ionising Radiation Effects.

MCMC Markov Chain Monte Carlo.

MD Muon Detector.

- MFC** Magnetic Flux Concentration.
- MIDAS** Met Office Integrated Data Archive System.
- MIP** Minimum Ionising Particle.
- MMF** Mean Magnetic Field.
- MOSWOC** Met Office Space Weather Operations Centre.
- MPV** Most Probable Value.
- NERC** Natural Environment Research Council.
- NIM** Nuclear Instrumentation Module.
- NM** Neutron Monitor.
- NMDB** Neutron Monitor Data Base.
- NOAA** National Oceanic and Atmospheric Administration.
- NUTS** No U-Turn Sampler.
- PCR** Primary Cosmic Ray.
- PMMA** Polymethylmethacrylate.
- PMT** Photo Multiplier Tube.
- PSD** Power Spectral Density.
- RM** Rotationally Modulated.
- RMS** Root Mean Square.
- RSS** Resonant Scattering Spectrometer.
- RV** Radial Velocity.
- SAPPHiRE** Simulation and Analysis Program Package for HiSPARC Research and Education.
- SB** Stochastic Background.
- SCR** Solar Cosmic Ray.
- SDO/AIA** Solar Dynamics Observatory Atmospheric Imaging Assembly.
- SDO/HMI** Solar Dynamics Observatory Helioseismic and Magnetic Imager.
- SEP** Solar Energetic Particle.
- SILSO** Sunspot Index and Long-term Solar Observations.

SMMF Solar Mean Magnetic Field.

SOHO/LASCO Solar and Heliospheric Observatory Large Angle and Spectrometric Coronagraph.

SOHO/MDI Solar and Heliospheric Observatory Michelson Doppler Imager.

SRB Solar Radio Burst.

SSN Sun Spot Number.

STFC Science and Technology Facilities Council.

SWPC Space Weather Prediction Center.

TTL Transistor-Transistor Logic.

UHECR Ultra-High-Energy Cosmic Ray.

UMR Unipolar Magnetic Region.

WDC World Data Center.

WSO Wilcox Solar Observatory.

1 Introduction

1.1 Solar Activity

1.1.1 Background

Solar activity varies periodically, with a duration of approximately 11 years, known as the solar cycle (Hathaway, 2015). The solar cycle is a magnetic effect and is driven by the Sun’s dynamo processes. It is well known that the 11-year solar activity cycle is in fact a 22-year cycle – the Hale cycle – which describes the alternating polarity and the full regeneration of the large-scale solar magnetic field (Hathaway, 2015; Charbonneau, 2020).

The Sun’s large-scale magnetic field alternates between a poloidal field and a toroidal field (Charbonneau, 2020), i.e. following the sequence:

$$P(+) \rightarrow T \rightarrow P(-) \rightarrow T \rightarrow P(+) \rightarrow T \rightarrow \dots, \quad (1.1)$$

where the P and T refer to poloidal and toroidal fields, respectively, (+) and (−) refer to the signs of the poloidal field polarity, and the different colours denote the ~ 11 -year cycles.

Starting with a poloidal field with positive polarity, $P(+)$, differential rotation causes shearing of the magnetic field and it wraps around the Sun, as shown in Figure 1.1 (Sanchez et al., 2014). As a result the magnetic field becomes concentrated at certain latitudes above and below the equator. This is named the Ω -effect and

is responsible for the creation of a toroidal field, T (Hathaway, 2015; Charbonneau, 2020).

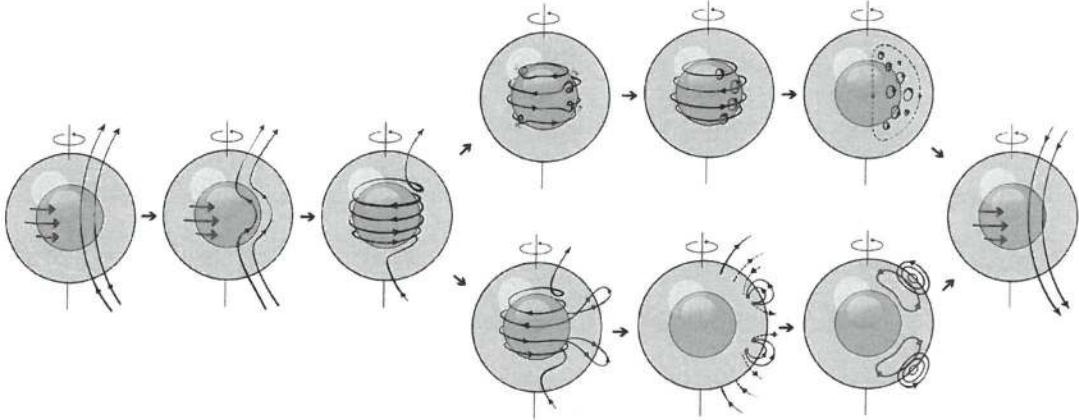


Figure 1.1: Schematic diagram of the main processes that drive the changes to the Sun’s magnetic field. Starting from a poloidal field on the left, it is shown how the Ω -effect generates a toroidal field. On the top we see a demonstration of the creation of small-scale poloidal magnetic fields, the α -effect. On the bottom we see a demonstration of the Babcock-Leighton mechanism, diffusion and meridional flows. Finally, on the right we see the recovery of a poloidal field with opposite polarity. Adapted from Sanchez et al. (2014).

The recovery of the poloidal field is not as straightforward. To recover a poloidal field with opposite polarity, $P(-)$, there are two main schools of thought, as shown in Figure 1.1 (Sanchez et al., 2014). One is based on the role of cyclonic turbulence and can be described by mean-field electrodynamics (Charbonneau, 2020). Due to the influence of the Coriolis force, helical motions lift and twist the magnetic field, known as the α -effect (Charbonneau, 2020). The degree of twist can produce sources of poloidal field and through the combination of many small, accumulating elements, we get the generation of a new poloidal field with opposite polarity (Charbonneau, 2020). Another mechanism is the Babcock-Leighton model (Babcock, 1961; Leighton, 1964) whereby tilted bipolar regions of magnetic field undergo diffusion and cross-equatorial reconnection. Through the action of diffusion and meridional flows, poloidal flux is advected to the poles which creates a new poloidal field with opposite polarity (Sheeley, 2005; Charbonneau, 2020). There is debate in the literature about which mechanism dominates the recovery of the poloidal field. These processes repeat with each cycle from poloidal-to-toroidal field and back

to a poloidal field with opposite polarity (i.e. $P(+)$ \rightarrow T \rightarrow $P(-)$), in ~ 11 years. Therefore, it takes ~ 22 years for the full magnetic cycle, the Hale cycle, to regenerate a poloidal field with the same polarity (i.e. $P(+)$ $\rightarrow \dots \rightarrow P(+)$).

The earliest observations of solar activity and the solar cycle were of sunspots, which date back to over 2000 years ago (Clark & Stephenson, 1978). Sunspots are dark regions of concentrated, active magnetic flux which cross the photosphere. The earliest observation of solar-cycle variability associated with sunspots was by Schwabe (1844), who found that annual observations of sunspot groups, over a duration 18 years from 1826–1844, showed a cyclic behaviour, with a period of about 10 years.

Following this discovery Wolf (1856, 1859) acquired daily observations of Sun Spot Numbers (SSNs) and extended the records as far back as 1749 (Hathaway, 2015). This used a method of estimating the number of individual sunspots and sunspots groups; these SSNs have since been used to monitor the solar activity cycle (Wolf, 1859, 1861). However, recently, a new, standard time-series of the SSN has been generated to reconcile various discrepancies between different observers (Clette et al., 2016; Clette & Lefvre, 2016): the International Sun Spot Number (ISSN), which is part of the Sunspot Index and Long-term Solar Observations (SILSO) activities of the Royal Observatory of Belgium (SILSO WDC, 2020). These SSNs are shown in Figure 1.2. Each activity cycle has unique characteristics and the SSNs have been used to measure the properties of the cycles, for sunspot observations dating back to the year ~ 1750 .

Observations of sunspots in the 17th Century allowed astronomers to accurately measure the rotation of the Sun's surface, which they observed to be slightly under four weeks (Casanovas, 1997; Casas et al., 2006; Luminet, 2017). These early observations also showed that sunspots did not appear at latitudes higher than $\sim 29^\circ/30^\circ$ (Casanovas, 1997). Later, it was confirmed that regions of strong surface magnetic activity were not distributed uniformly over the solar surface and also that the Sun

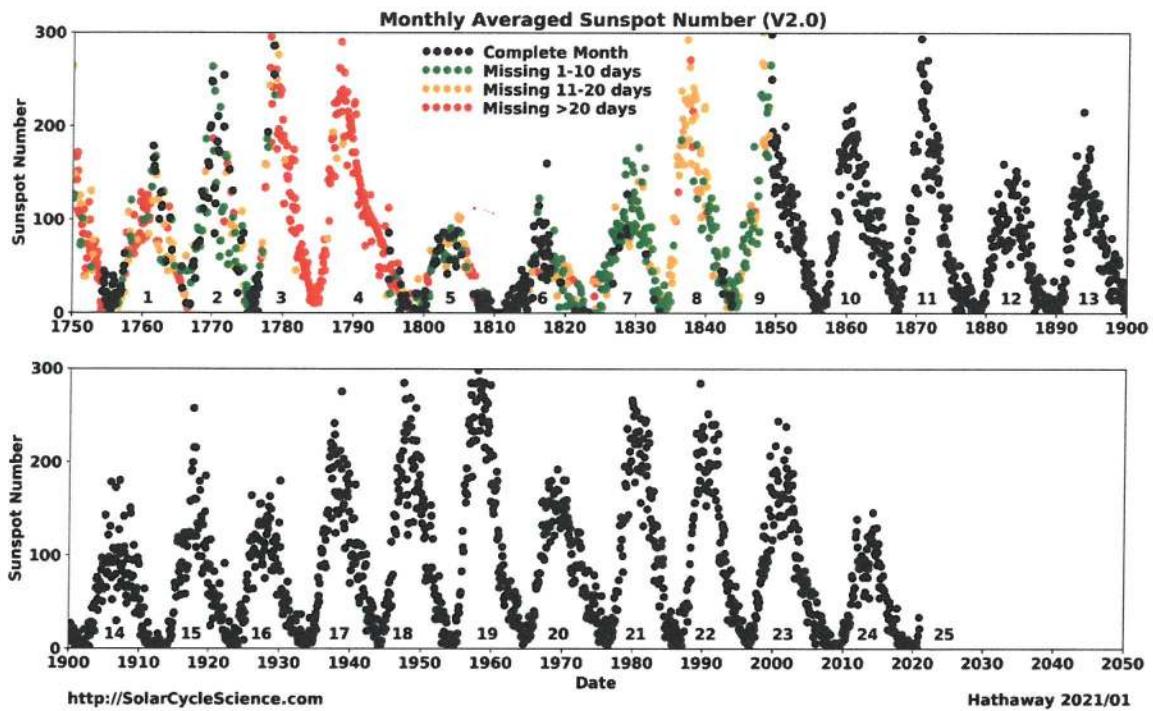


Figure 1.2: Monthly averaged sunspot number since 1750 (Hathaway & Upton, 2017). Black points indicate a full month of data, and then a traffic light system is in place to represent data quality, i.e. green for a near-complete month of data, amber for near 50% duty cycle, and red for a very low data duty cycle.

*just describe this as in legend,
e.g. missing 1-10 days etc.*

exhibited a latitude-dependent differential surface rotation (Lee, 1858).

As the solar cycle evolves, the latitudes displaying sunspots decreases and the typical latitude of spots slowly drifts towards the equator, with a zone of avoidance near the equator (Hathaway, 2015). This behaviour was first noticed by Carrington (1863) and is known as Spörer's Law of Zones, illustrated by the "Butterfly Diagram" (Maunder, 1903, 1904), such as the one shown in Figure 1.3.

In the first half of the 20th Century, it was discovered that sunspot groups in both hemispheres were tilted with respect to the Sun's equator, such that the leading spots (i.e. leading in the sense of the direction of rotation) exist closer to the equator than the following spots. This was first published by Hale et al. (1919) but later defined as Joy's law. Furthermore, the degree of tilt varies with latitude, with a larger tilt at higher latitudes, and also with the solar cycle (Hathaway, 2015). In addition, sunspots groups have opposite polarities in the Northern and Southern hemispheres, and the polarity changes from cycle-to-cycle. This effect is known as

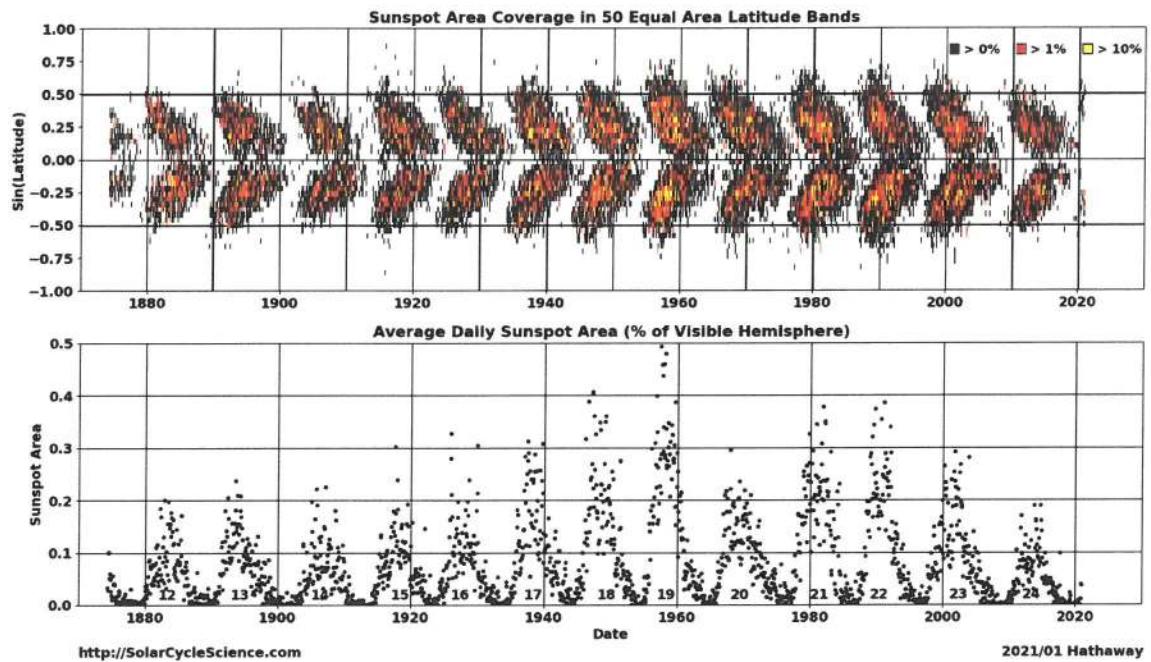


Figure 1.3: Top panel: shows the latitudinal distribution of sunspots over time (also known as a “Butterfly Diagram”) and the colour of the points depict the area of the disc. Bottom panel: shows the average daily area of sunspots on the visible solar disc. Taken from [Hathaway & Upton \(2017\)](#).

Hale’s Polarity Law (Hale & Nicholson, 1925).

With the invention of the magnetograph (Babcock, 1953), the Sun’s magnetic field was probed further and it was found that sunspots were a characteristic of a larger phenomenon. It was discovered that sunspots are a small part of larger Active Regions (ARs) of magnetic field, which may be both Bipolar Magnetic Regions (BMRs) or Unipolar Magnetic Regions (UMRs) (Babcock & Babcock, 1955).

Joy’s law implies that the tilt of BMRs systematically places the leading-polarity-flux at a lower latitude than the following-polarity-flux (Hathaway, 2015). It is believed that there is a cancellation of leading-polarity flux on the opposite hemisphere near the equator (Dasi-Espuig et al., 2010).

Recently, In very modern literature, the definition of the solar activity cycle has started to deviate away from counting sunspots and instead tracking the magnetic activity bands over the full, 22-year, Hale cycle (Leamon et al., 2020). The existence of a “Terminator” has been found, which marks the hand-over from one solar cycle to the

next (McIntosh et al., 2014, 2019). The terminator is defined as the abrupt annihilation or cancellation event of the oppositely polarised magnetic activity bands at the solar equator (McIntosh et al., 2019). By employing the Hilbert transform on different proxies of the solar cycle, Leamon et al. (2020) demonstrated a robust method to identify and predict the signature of terminators. The method has since been used to map severe space weather events to a “solar cycle phase clock” (Chapman et al., 2020) and forecast the properties of Solar Cycle 25 (McIntosh et al., 2020). There exists some uncertainty about terminators in the solar physics community; the evidence presented so far suggests this method is robust, but will ultimately be tested with the evolution of Solar Cycle 25.

As discussed, the solar activity cycle exhibits a clear periodicity of approximately 11 years; however, there exist other, short-term periodicities in the solar activity cycle (see Hathaway (2015) for a review). The most prominent other feature in the solar activity cycle is an \sim 2-year periodicity which manifests as a double peak in the maximum of SSN, known as the Gnevyshev gap (Gnevyshev, 1963, 1967). It was thought that this phenomenon was due to a superposition of sunspots from both hemispheres, slightly out of phase, but it was later confirmed that the Gnevyshev gap occurs in each hemisphere separately (Norton & Gallagher, 2009). The cause of the Gnevyshev effect remains an open question (Hathaway, 2015).

Finally, it has been suggested that long-term variations exist in the solar activity cycle, which consist of grand maxima and grand minima, which have led to the events such as the Maunder minimum (circa 1650–1700) and the Dalton minimum (circa 1790–1820). However, there is little evidence to support this long-term, cyclic variation (Hathaway, 2015).

1.1.2 Features of Solar Activity

There are many features that can be observed across the electromagnetic spectrum as a result of solar activity. In addition, with the advent of the magnetograph, solar

activity can also be observed with direct measurements of the Sun's magnetic field, at different spatial scales. Many of these features are used widely in the literature as proxies of the solar activity cycle. Here we discuss these features and their properties.

Active Regions

An Active Region (AR) has been defined as all *observable* magnetic phenomena that appear preceding, during, and following the birth of sunspots, including electromagnetic and Solar Energetic Particle (SEP) emission (Kiepenheuer, 1968). However, more generally, they appear when strands of magnetic flux emerge into the visible atmosphere from the solar interior.

Recently van Driel-Gesztelyi & Green (2015) ~~recently~~ proposed a new and more accurate definition of ARs:

the totality of observable phenomena in a 3D volume represented by the extension of the magnetic field from the photosphere to the corona, revealed by emissions over a wide range of wavelengths accompanying and following the emergence of strong twisted magnetic flux (i.e. kG).

This definition does not hinge on the appearance of a sunspot, which is important as not all ARs produce spots (van Driel-Gesztelyi & Green, 2015), and clearly defines their conditions and features. ARs comprise ~~of~~ features in the photosphere, chromosphere, and corona; they typically appear as bipolar regions, obeying Hale's law, Joy's law, and Spörer's Law, and display a wide range of sizes and lifetimes. In addition, ARs are typically confined to the specific 'active' bands of latitude, between $\pm 40^\circ$ (Harvey, 2001).

The ARs which produce sunspots are typically 10–100 Mm in size and have lifetimes on the order of several months (Zwaan, 1981; Schrijver & Harvey, 1994; Canfield, 2001; van Driel-Gesztelyi & Green, 2015). However, it has been observed that ARs exist with lifetimes in the range of days–months (Schrijver & Zwaan, 2008). It is generally accepted that the lifetimes of ARs depend approximately linearly on

their size and strength (Canfield, 2001; Schrijver & Zwaan, 2008). The end of an AR is determined by the dispersal of the magnetic flux by convective motion, differential rotation, meridional flows, and magnetic reconnection (Canfield, 2001).

There are several other features that arise out of ARs, such as: faculae, plages, and networks. Faculae are bright spots visible in the photosphere that appear bright at the limb but are nearly invisible at disc centre. They arise due to small magnetic flux tubes creating pores in the photosphere (Solovev et al., 2019). If the flux tubes are small enough, they are heated by horizontal radiation transfer and the walls appear brighter than the surrounding area, which is known as the ‘hot-wall’ model (Spruit, 1976; Keller et al., 2004). Plages are bright regions in the solar chromosphere and are the chromospheric counterparts of faculae (Pillet et al., 1997). Networks are a product of the decaying remnants of ARs, the merging and cancelling of mixed polarity fields in and around ARs, and ephemeral regions – often referred to as the ‘salt and pepper’ field (Martin, 1988).

Ephemeral Regions

A short-lived class of Active Regions (ARs) exists, known as Ephemeral Regions (ERs) due to their short lifetimes. ERs have sizes typically less than 20 Mm, lifetimes of a few hours, and produce no sunspots (Harvey, 2001).

ERs emerge rapidly and it is possible that thousands of ERs emerge over the entire solar surface within a 24-hour period (Harvey, 2001). There has been some disagreement in the literature on the temporal evolution of ERs with contradictory statements regarding whether the occurrence of ERs follows the activity cycle (see Harvey & Zwaan, 1993; Hagenaar, 2001; Vieira & Solanki, 2010). However, some studies show compelling evidence that ERs are correlated with the solar activity cycle (see e.g. Vieira & Solanki, 2010; Chaplin et al., 2019, and references therein). It was also shown that ERs follow the solar activity cycle but with a slight shift in phase compared to the sunspot minimum (Harvey & Martin, 1973; Martin & Harvey, 1979).

Unlike ARs, ERs are spatially located all over the solar disc (Harvey, 2001). Furthermore ERs at high latitudes have been interpreted to be the first emergence of bipolar regions associated with the onset of a new solar cycle. Despite these differences, it has been reported that it is difficult to differentiate between large ERs and small ARs (Harvey, 2001).

Sunspots

As discussed above, sunspots have been used for centuries as a proxy for solar activity. Sunspots are dark regions on the disc of concentrated magnetic field. The convective heat flow is disrupted due to the magnetic field, therefore leaving the spots cooler than the surrounding photosphere (Solanki, 2003; Hathaway, 2015). Sunspots can sometimes appear in isolation; however, sunspots are generally found in groups, obeying Hale's law, Joy's law, and Spörer's Law. Each sunspot is characterised by a dark, central core, called the *umbra*, and a halo, called the *penumbra* (Howard, 2001; Solanki, 2003). Sunspot groups are an important photometric counterpart of Active Regions (ARs) as the locations of extreme solar activity such as flares and Coronal Mass Ejections (CMEs).

The sizes of sunspots can vary substantially from diameters of around 2000 km, for the smallest sunspots, up to 60 Mm or more (Howard, 2001; Solanki, 2003). The lifetimes of sunspots range from minutes to months (Howard, 2001; Solanki, 2003) and Howard (2001) states that more than half of sunspots have lifetimes < 2 days, 90% of sunspots have lifetimes < 11 days, and it is rare for sunspots to have lifetimes of more than a few months. The lifetime increases linearly with maximum size, following the Gnevyshev-Waldmeier rule (Gnevyshev, 1938; Waldmeier, 1955). The largest spots are generally the survivors of large sunspot groups, appearing as a single spot at least in the later stage of their lifetime (Howard, 2001).

The equatorward drift of sunspots is a well-known characteristic of their evolution during the solar cycle. Sunspots emerge in active latitudes at around $\sim 25-30^\circ$, and as the cycle progresses the mean latitude of the spots in each hemisphere steadily

decreases. The number of sunspots varies with time with more spots observed during solar maximum and fewer observed at solar minimum (see Fig. 1.2). As such, the sunspot number is the most commonly used proxy for solar activity (Hathaway, 2015) and the International Sun Spot Number (ISSN) is the most widely used data set due to the length of the available record (Hathaway, 2015).

Sun Spot Numbers (SSNs) are preserved and disseminated by the World Data Center (WDC) Sunspot Index and Long-term Solar Observations (SILSO) of the Royal Observatory of Belgium (SILSO WDC, 2020) and are provided as the daily SSN, monthly averages, yearly averages, and the box-car smoothed SSN. The standard smoothing is a 13-month Central Moving Average (CMA) and the solar cycle maxima and minima are usually defined in terms of the smoothed SSN (Hathaway, 2015).

10.7 cm Flux

The 10.7 cm solar flux ($F_{10.7}$) is the disc-integrated radio emission from the Sun at a wavelength of 10.7 cm (in a 100 MHz-wide band centred on 2800 MHz) (Tapping & Charrois, 1994; Tapping, 2013). It is one of the most widely used indices of solar activity, alongside the SSN, and the $F_{10.7}$ measure of solar activity has advantages over the SSN in that it is completely objective (Hathaway, 2015).

The $F_{10.7}$ comprises of Bremsstrahlung emission from the chromosphere and corona as well as over sunspots, where the magnetic fields are sufficiently strong to produce bright, compact sources from thermal, free-free electron gyro-resonances (Tapping & DeTracey, 1990; Tapping, 2013). However, the dominant component of the $F_{10.7}$ originates in the low corona (Tapping & DeTracey, 1990).

The first observations were made in 1946 and have been recorded consistently since 1947 (Covington, 1969; Tapping, 2013). The data can be acquired from the Laboratory for Atmospheric and Space Physics at University of Colorado (LISIRD, 2019).

Solar Eruptions

Solar flares are energetic, eruptive events which occur in the region of Active Regions (ARs) and complex sunspot groups (Hathaway, 2015). The first reported solar flare was in 1859, the largest known flare to-date, the ^{so called} "Carrington" event (Carrington, 1859).

Solar flares follow the solar activity cycle, with more flares occurring at solar maximum, which is because during these times there are more ARs on the Sun which tend to be flare-producing (Gopalswamy, 2010; Hathaway, 2015). However, there is a general propensity for more flares to occur during the declining phase of a sunspot cycle (Hathaway, 2015).

Coronal Mass Ejections (CMEs) are manifestations of concentrated solar activity and are associated with the significant release of plasma and accompanying magnetic field ejected from the corona. CMEs are often associated with solar flares but can also occur in isolation, in the absence of a flare (Hathaway, 2015).

CMEs were first discovered in the early 1970s via spacecraft observations and have since routinely been observed (Hathaway, 2015). A database of the properties of all CMEs observed since 1996 by the Solar and Heliospheric Observatory Large Angle and Spectrometric Coronagraph (SOHO/LASCO) instrument exists in the SOHO/LASCO catalogue¹ (Yashiro et al., 2004; Gopalswamy et al., 2009).

Similar to flares, the frequency of CMEs follow the solar activity cycle, with more CMEs occurring at solar maximum (Gopalswamy, 2010) and is therefore correlated with the ~~Sun Spot~~ Number (SSN) (Webb & Howard, 1994). In addition, physical properties of CMEs such as average speed, size, and the location of ejection from the Sun, etc. also vary with the solar activity cycle (Yashiro et al., 2004).

¹https://cdaw.gsfc.nasa.gov/CME_list/

Total Magnetic Flux

Individual elements of the magnetic field contribute to the large-scale field and determine the Sun's global magnetic dipole moment and the strength of the field in interplanetary space (Lockwood et al., 1999; Solanki et al., 2000). The total, unsigned, magnetic flux is the amount of flux leaving the Sun and entering the heliosphere, i.e. the combination of AR, Ephemeral Region (ER), and open magnetic flux (Lockwood et al., 1999). The strength of the total magnetic flux and its main components has been shown to vary with the solar activity cycle (Solanki et al., 2002; Vieira & Solanki, 2010; Chaplin et al., 2019). Vieira & Solanki (2010) showed that the respective contributions to the total flux during maximum activity are approximately in the ratio 0.1:0.8:0.1, for ARs, ERs, and open flux, while the ratios during minimum activity are approximately 0.5:0.4:0.1, for ARs, ERs, and open flux.

The Solar Mean Magnetic Field

The Solar Mean Magnetic Field (SMMF) is the mean Line Of Sight (LOS), signed magnetic field when observing the Sun-as-a-star (Scherrer et al., 1977a,b; Garca et al., 1999). The SMMF is surprisingly a non-zero measurement of the imbalance of opposite magnetic flux polarities observed on the full, visible disc of the Sun (Svalgaard et al., 1975), and its amplitude is observed to vary with the solar activity cycle.

The literature on SMMF observations spans several decades; however, the origin of the SMMF remains an open question in solar physics. We give a comprehensive review of the morphology of the SMMF in Section 5.1, since the analysis and interpretation of SMMF data collected by the Birmingham Solar Oscillations Network (BiSON) is presented in detail in that chapter.

1.2 Space Weather

1.2.1 Background

Space weather is defined as (Cannon & Royal Academy of Engineering (Great Britain), 2013):

variations in the Sun, solar wind, magnetosphere, ionosphere, and thermosphere, which can influence the performance and reliability of a variety of space-borne and ground-based technological systems and can also endanger human health and safety.

I think this
is actually
the US
definition from
1995.

Space weather phenomena have been observed over hundreds of years, mainly through observations of the aurorae, but its impacts are slowly becoming more tangible in modern civilisation, as we grow reliant on electronics (Beggan et al., 2018).

There are two main sources of space weather: (i) those that are solar in nature and (ii) those whose origins are external to the solar system but penetrate into the heliosphere. Space weather manifests itself broadly in three ways:

1. Electromagnetic radiation, which is generally linked with an enhancement in the output of the Sun's spectrum.
2. Magnetic fields/plasma, which can cause disturbances in the Interplanetary Magnetic Field (IMF), solar wind, and Earth's magnetosphere.
3. Energetic charged particles, which refer to ionising charged particles and ions.

The nature of their arrival and the resulting impact of these space weather phenomena at Earth depends on their type and energy. Space weather is the resultant product from solar storms – magnetic disturbances on the Sun leading to large bursts of energy release and short-term heliospheric effects – and the general chronology of

events was outlined by Cannon & Royal Academy of Engineering (Great Britain) (2013):

1. Storms begin with the evolution of one or more complex sunspot groups and Active Regions (ARs) on the solar surface.
2. Within ARs, one or more solar flares occur and the electromagnetic radiation is detected on Earth within approximately 8 minutes.
3. Solar Energetic Particles (SEPs) are released and are measured at Earth, both using satellites and ground-based detectors, within approximately 15 minutes. SEPs continue to arrive over a period of several hours–days.
4. A Coronal Mass Ejection (CME) occurs and propagates outwards, arriving at a distance of 1 AU within ~15–72 hours. The impact on Earth depends on the CME speed, how close it passes to Earth, and the orientation of the magnetic fields, with southward magnetic field generating the most severe geomagnetic storms because of its interference with Earth's northward magnetic field resulting in reconnection at the magnetopause.

This isn't quite true, it depends. See Cliver & Svalgaard 2004.

The largest documented space weather event since modern records began occurred in 1859, the solar super storm known as the “Carrington event” (Carrington, 1859). The available measurements of this event are limited to geomagnetic field perturbations, as well as eye witness accounts of solar brightening and aurorae (Cannon & Royal Academy of Engineering (Great Britain), 2013). However, recently cosmonuclide measurements from ice cores have been used to learn about the properties of the Carrington event (Riley, 2012). *Like what?*

Since the beginning of the space age there have been no other super storms, however there have been large storms that have affected the infrastructure and caused a significant economic impact. The consensus is that another storm of the Carrington event level is inevitable and could significantly impact society. There

Quite a lot of work has been done recently on improving the estimates for this. Specifically for flares I have a paper (Elvidge & Angling 2018), or more generally Love 2020 or 2021. On the whole 100 yrs seems more likely.

is a view that a Carrington-like event may occur again in a period of 250 years with a confidence of ~95% and within a period of 50 years with a confidence of ~50% (Cannon & Royal Academy of Engineering (Great Britain), 2013); however, it is stressed that these figures should be interpreted with care. It was suggested by Riley (2012) that a Carrington-like event may occur with ~12% probability between 2012–2022 and later in 2012 a large storm occurred, missing Earth, but it strongly interacted with the STEREO-A satellite (Russell et al., 2013). This near miss highlights that Carrington-level events are a real threat to society and that we need a method of predicting their occurrence, arrival, and impact.

1.2.2 Impacts of Space Weather

Space weather is an increasingly tangible threat to modern infrastructure and society, due to the increasing reliance on electronic technology. In 2011 space weather was added to the UK National Risk Assessment for the first time, and the subsequent National Risk Register in 2012 (BIS, 2015) where it has remained to-date (Government, 2020). At the time of writing this thesis, space weather risk was rated as a medium severity/high likelihood risk, at the same level as emerging infectious diseases, poor air quality, and heatwaves (see Figure 1.4) (Cabinet Office, 2017).

An alarming aspect of Figure 1.4 is that the likelihood of space weather events occurring in 5 years from 2017 was rated the same as pandemic influenza. Finishing this Ph.D. during a global pandemic highlights the importance of taking this risk register very seriously. We should learn from the global response to the COVID-19 pandemic and ensure that the world has a resolute plan to deal with the occurrence of a severe space weather event.

There are many ways that technological systems are impacted by space weather, both on or above ground, and Figure 1.5 displays many of the key impacts that we know of (Beggan et al., 2018).

Mention love Hapgood et al. 2021 which is a review of the situation from a UK viewpoint. This also mentions COVID-19.

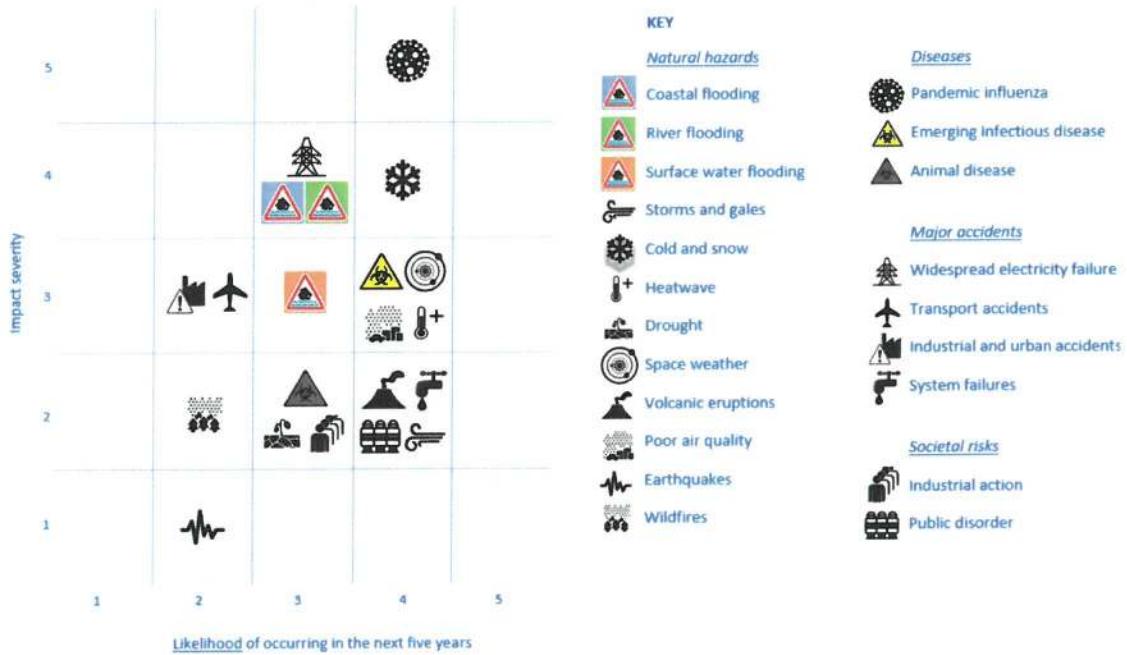


Figure 1.4: UK National Risk Register for hazards, diseases, accidents, and societal risks showing space weather as a medium-high risk (Cabinet Office, 2017)



Figure 1.5: The sources and effects of space weather. Impacts are shown including loss of telecommunications and GNSS, increased radiation levels, and ground induced currents (ESA/Science Office, CC BY-SA 3.0 IGO)

As far back as October 1841, it was reported that a solar storm was responsible for railway disruptions around Exeter, due to magnetic disturbances making it impossible to ascertain whether the onward line was clear, leading to delays (Nature, 1871). It is possible for space weather events to induce geomagnetic storms that can cause damaging Ground Induced Currents (GICs) within large power grids, causing them to fail. Two such famous cases of GIC grid failures were in Quebec, Canada in 1989 which resulted in the failure of the Quebec-Hydro grid for 9 hours, and the city-wide black-out in Malmö, Sweden, during the Halloween storm in 2003 (Viljanen, 2011; Beggan et al., 2018).

It was documented that in May 1967, the U.S. Air Force was close to engaging in conflict with its enemies – during those politically tense times – due to the misinterpretation of the effects of space weather (Knipp et al., 2016). Solar Radio Burst (SRB) induced radio frequency interference was initially interpreted as surveillance jamming, an act of war. Fortunately, the U.S. had begun monitoring the space environment and were able determine that space weather was really the cause of the disruption, hence avoiding further conflict (Knipp et al., 2016). Furthermore, in more modern times, scintillation effects induced in the ionosphere affect Global Navigation Satellite System (GNSS) communications which has a large societal effect due to our reliance on GNSS (Cannon & Royal Academy of Engineering (Great Britain), 2013) as shown in Figure 1.5.

Solar storms are also responsible for creating sudden increases in ionising radiation which – at typical flight altitudes – can lead to the risk of malfunctions in aircraft microelectronic systems and unquantified radiation doses to passengers and crew (Cannon & Royal Academy of Engineering (Great Britain), 2013). In orbit, these conditions threaten the operation of satellites and the safety of manned space endeavours which is of particular concern with the current ambitions to return to the Moon and voyage to Mars.

There are even concerns that major storms could cause radiation increases at

the Earth's surface, Ground Level Enhancements (GLEs), which may cause malfunctions in microelectronics that are likely to be of increasing concern in the design of safety-critical systems (Cannon & Royal Academy of Engineering (Great Britain), 2013).

I think newer estimates have revised this down. See Eastwood, 2017.

Finally, a cost analysis estimated a present-day total U.S. economic cost for a super storm on the scale of the 1859 Carrington event (Homeier & Wei, 2013). The cost estimated is heavily dependent on the duration of outages, the damages caused, and the availability of spare parts for repair; however they estimated the impact on the U.S. economy to be at around \$0.6–2.6 trillion (Homeier & Wei, 2013). These figures show the large scale impact that space weather can have on the economy and that mitigation techniques to reduce this cost are of imperative necessity.

Due to the many ways that space weather can impact civilisation, and that it is predicted that there is a significant probably of the re-occurrence of large solar storms, it is easy to understand why space weather forecasting is becoming increasingly more necessary as a mitigation technique.

The U.S. National Oceanic and Atmospheric Administration (NOAA) is the leading global space weather forecasting agency. The NOAA Space Weather Prediction Center (SWPC) gathers data in real-time to describe the conditions of the Sun, heliosphere, magnetosphere, and ionosphere to understand the environment within the heliosphere and on Earth. With these data, the SWPC produces forecasts, warnings, and alerts available to inform anyone concerned and affected by space weather (NOAA, 2018).

Following the addition of space weather to the UK risk register, the UK Met Office Space Weather Operations Centre (MOSWOC) opened in 2014 (BIS, 2015). MOSWOC is mandated to produce daily space weather forecasts and is therefore developing a forecasting infrastructure using ground-based and satellite instrumentation to monitor space weather events. In addition, scientific research at MOSWOC is carried out to better understand the physical processes involved in space weather

phenomena to improve forecasting accuracy and lead-times; current forecasting enables prediction of CMEs impacting Earth to within only plus or minus six hours at best (MetOffice, 2013).

Forecasting and prediction is one aspect of the response to severe space weather events. We must also learn from the global response to the COVID-19 pandemic and ensure that upon the occurrence of a severe space weather event, suitable pre-planning has been performed and a sufficient contingent action is planned.

Another lesson from COVID I think is to have a good comms strategy.

1.3 Cosmic Rays

1.3.1 Background

Cosmic Rays (CRs) are charged particles and atomic nuclei with energies spanning from 1 keV up to around 10^{21} eV, that encroach upon the Earth from all directions (Giacalone, 2010). It is understood that CRs are composed of $\sim 99\%$ of atomic nuclei and $\sim 1\%$ electrons (Gaisser et al., 2016); of the atomic nuclei $\sim 87\%$ are protons, $\sim 12\%$ are α -particles, and a smaller contribution of $\sim 1\%$ are heavier nuclei (Grupen, 2005; Dunai, 2010; Particle Data Group et al., 2020). CRs mainly originate from outside the solar system, known as Galactic Cosmic Rays (GCRs) (Particle Data Group et al., 2020). These GCRs mostly come from within the Milky Way, although they are also expected to emanate from extra-galactic sources, in particular for CRs with energies above 10^{18} eV (Aab et al., 2017). Incoming low-energy CRs ($\lesssim 1$ GeV/nucleon) are modulated by the solar wind, which decelerates GCRs and can even forbid lower-energy GCRs from the inner solar system (Grupen, 2005). Consequently, there exists a strong anti-correlation between solar activity and the GCRs flux (Particle Data Group et al., 2020).

Cosmic rays produced within the heliosphere are mostly of solar origin, known as Solar Cosmic Rays (SCRs) or Solar Energetic Particles (SEPs). These SCRs are generally of a lower energy than GCRs and may be accelerated in the solar wind, by interplanetary shocks, or in solar eruptions (e.g. solar flares) (Giacalone,

I know this is fussy, but you should either have a space or not (my preference) consistently.
Do a ctrl-F and replace on the whole document.

2010). SCRs have typical energies on the order of magnitude of $\sim 10^1$ keV to 1 GeV (Chilingarian, 2003; Bruno et al., 2018). Therefore, Primary Cosmic Rays (PCRs) associated with space weather events are generally of much lower energy than the background of GCRs.

The intensity spectrum of PCRs in the energy range from 10^9 eV to $\sim 10^{14}$ eV is given approximately by:

$$I_N(E) = \frac{dN}{dE} \approx 1.8 \times 10^4 (E/1\text{ GeV})^{-\alpha} \frac{\text{nucleons}}{\text{m}^2 \text{s sr GeV}}, \quad (1.2)$$

where E is the energy per nucleon (including rest mass energy) in GeV and $\alpha = 2.7$ is the differential spectral index of the cosmic-ray flux (Particle Data Group et al., 2020).

Figure 1.6 shows a graphical representation of the CR energy spectrum described by equation (1.2). It shows the flux for a number of CR species, over the energy range 10^{-1} to 10^6 GeV/nucleus, measured by several different experiments.

Beyond the x-axis range in Figure 1.6, the spectrum ‘knee’ occurs (i.e. in the range $\sim 10^{15}$ to 10^{16} eV) where the spectral index is thought to increase to ~ 3 (Particle Data Group et al., 2020). At even higher energies, in the region of the spectrum ‘ankle’ ($\sim 10^{18.5}$ eV) the spectral index reduces and the spectrum becomes less steep. This is in the regime of Ultra-High-Energy Cosmic Rays (UHECRs) and the interaction between GCRs and photons of the Cosmic Microwave Background (CMB) sets an upper limit on their energy, the Greisen-Zatsepin-Kuzmin (GZK) limit (Particle Data Group et al., 2020). The GZK limit implies that CRs with energies exceeding $\sim 5 \times 10^{19}$ eV must have originated from distances within a horizon of ~ 50 Mpc, as otherwise their energy would have been reduced by the GZK effect (Particle Data Group et al., 2020).

Propagation of CRs through magnetic fields depends on their gyroradius or Larmor radius (Particle Data Group et al., 2020). Therefore, a common description of CRs uses a property called the *magnetic rigidity* which is defined by:

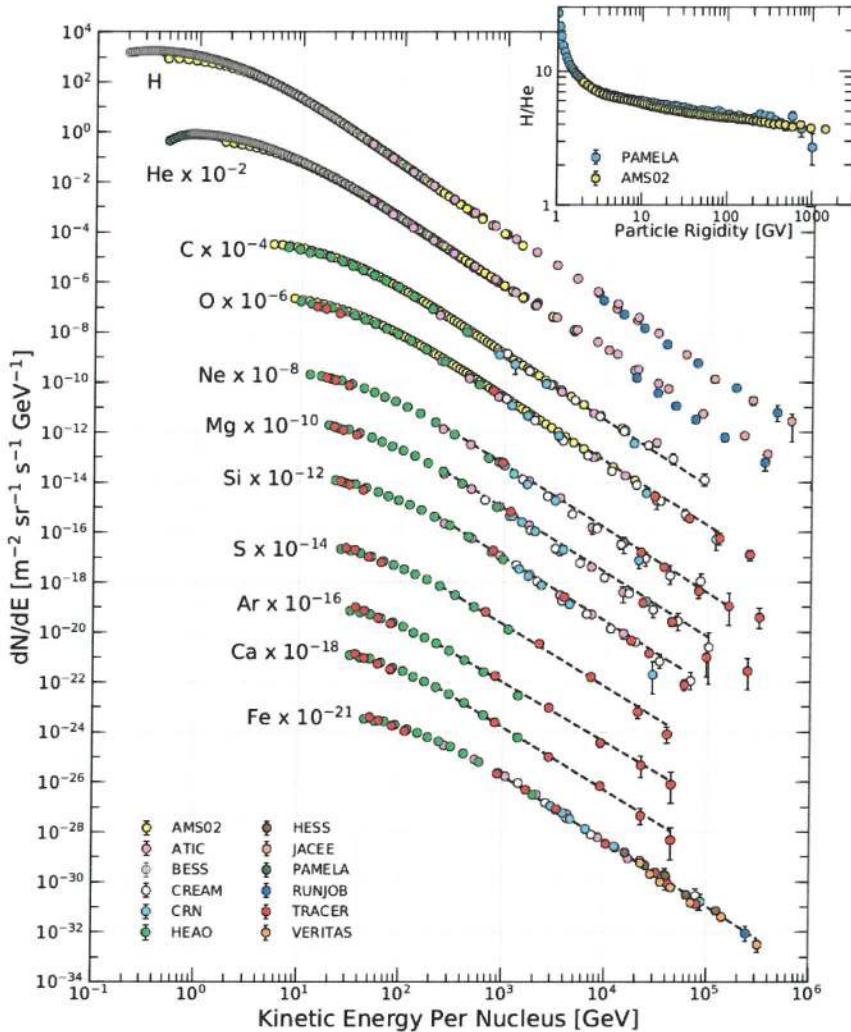


Figure 1.6: Cosmic ray differential energy spectrum using data measured by several experiments. The inset shows the H/He ratio at constant rigidity (Particle Data Group et al., 2020)

$$R = r_L B c = \frac{p c}{Z e}, \quad (1.3)$$

where r_L is the Larmor radius, B is the magnetic field strength, c is the speed of light, p is the particle's momentum, Z is the atomic number, and e is the electron charge. The magnetic rigidity has units of Volts (V), or usually due to a large magnitude, Gigavolts (GV). The rigidity is used to describe CRs as particles with different charges and masses have the same dynamics in a magnetic field if they have the same rigidity (Particle Data Group et al., 2020).

1.3.2 Cosmic Rays in the Atmosphere

CRs in the interstellar medium traverse a very low-density medium, but experience a much denser environment when they reach the atmosphere. The typical nucleon mean free path (measured in units of g cm^{-2}) of protons in the atmosphere is approximately 90 g cm^{-2} , which means the first interactions of CRs occur in the upper layers of the atmosphere, at a height of $\sim 15\text{--}20 \text{ km}$ (Grupen, 2005).

The PCRs will predominantly interact with the atmospheric nuclei via strong interactions (Grupen, 2005). When PCRs interact with atmospheric nuclei, the interaction leads to the production of a cascade of secondary particles. The secondary particles can also undergo interaction or decay, producing tertiary particles, and the process continues until the energies of all particles are insufficient to create new particles. If a concentrated, large number of secondary particles reach ground-level, the cascade of particles is called an Air Shower (AS), or an Extensive Air Shower (EAS) for extremely high numbers of secondary particles, which can have a footprint area of several km^2 (Fokkema, 2012; van Dam et al., 2020b). The AS is often described as being a cone, with the base being the shower front and the apex being the primary CR.

Hadronic cascade components (or the “hard component” of ASs) are produced by CR protons and nuclei interacting with atmospheric nuclei. This process typically produces lower energy protons, neutrons, pions, and kaons. In this thesis we are mostly interested in the muonic air shower development (see Section 1.4), so here we will focus on the development of the mesons in the ASs, as they predominantly produce muons. In Table 1.1, the most likely modes of decay for air shower mesons are shown with the branching ratio for each mode. In addition, the table shows the most likely decay modes of muons.

Due to the short lifetimes of pions and kaons, $\sim 26 \text{ ns}$ and $\sim 12 \text{ ns}$, respectively (Particle Data Group et al., 2020), and depending on energy, they decay during their journey to ground-level. It is shown in Table 1.1 that the most probable decay modes

Table 1.1: Most prominent decay modes of the mesonic components of CR air showers and of muons. Note: K^- modes are charge conjugates of the decay modes below (Particle Data Group et al., 2020)

Decay mode	Branching ratio
$\pi^+ \rightarrow \mu^+ + \nu_\mu$	$99.98770 \pm 0.00004\%$
$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$	$99.98770 \pm 0.00004\%$
$\pi^0 \rightarrow \gamma + \gamma$	$98.823 \pm 0.034\%$
$\pi^0 \rightarrow e^+ + e^- + \gamma$	$1.174 \pm 0.035\%$
$K^+ \rightarrow \mu^+ + \nu_\mu$	$63.56 \pm 0.11\%$
$K^+ \rightarrow \pi^+ + \pi^0$	$20.67 \pm 0.08\%$
$K^+ \rightarrow \pi^+ + \pi^+ + \pi^-$	$5.583 \pm 0.024\%$
$K^+ \rightarrow \pi^0 + e^+ + \nu_e$	$5.07 \pm 0.04\%$
$K^+ \rightarrow \pi^0 + \mu^+ + \nu_\mu$	$3.352 \pm 0.033\%$
$K^+ \rightarrow \pi^+ + \pi^0 + \pi^0$	$1.760 \pm 0.023\%$
$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$	$\sim 100\%$
$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$	$\sim 100\%$

involve the production of muons. Muons are the most abundant charged particles at ground level (Particle Data Group et al., 2020); on average, the ground level muon flux is on the order of $\sim 70 \text{ m}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ (Cecchini & Sioli, 2000; Blackmore et al., 2015; Pereira et al., 2021; Particle Data Group et al., 2020). Most muons are produced high in the atmosphere (~ 15 km) and lose about 2 GeV to ionization before reaching the ground and the mean energy of muons at the ground is ~ 4 GeV (Particle Data Group et al., 2020).

In addition to the hadronic component of an air shower, electron and photon constituents of cascades are called the electromagnetic component (or “soft component”). These are typically initiated by electrons and photons under the processes of Bremsstrahlung (Grupen, 2005),

$$e \rightarrow e + \gamma, \quad (1.4)$$

or pair production (Grupen, 2005),

$$\gamma \rightarrow e^- + e^+. \quad (1.5)$$

Ionisation losses mean that electrons and positrons lose energy rapidly until they either annihilate or recombine with nuclei, and photons lose their energy by being either absorbed in scattering and/or the photoelectric effect. Therefore, most of the electrons, positrons, and photons observed at ground level are produced from the decaying hadronic AS component and muon decay is the dominant source of low-energy electrons at ground level (Particle Data Group et al., 2020).

Finally, there is a minimum rigidity cut-off which implies that the energy of any PCR must exceed a minimum energy to be able to initiate an AS or particle cascade and be measured at ground-level. This limit is dependent on the depth of the atmosphere above the detector, but is greatest at sea-level and decreases with increasing altitude. The minimum energy to be measured at sea-level is approx. 430 MeV/nucleon (Dorman, 2004a,b; Poluianov et al., 2017).

1.3.3 Cosmic Ray Detectors

To observe CRs there are many types of usable detectors, both ground-based and space-based (Schrijver & Siscoe, 2010); in this thesis we are mostly concerned with ground-based detectors of the AS hadronic component. The most common type of ground-based CR detectors are Neutron Monitors (NMs) and Muon Detectors (MDs) which indirectly measure CR particles through the secondary particles produced in CR cascades. These two types of detector probe different energy ranges; NMs generally observe PCRs with energies \sim 1–10 GeV and above, while MDs typically observe higher energy PCRs with energies on the order of \gtrsim 10 GeV (Kuwabara et al., 2006a; Rockenbach et al., 2014).

Neutron Monitors

The neutron monitor, invented by Simpson (1948), has been extensively used for CR observations of the space environment (Clem & Dorman, 2000). The NM is an example of an ionisation detector whereby energetic neutrons encounter a nucleus within a gas, producing charged, secondary particles which in turn ionise the surrounding gas (Gloeckler, 2010).

The original “IGY” NM design made use of a paraffin reflector to trap slow neutrons within the detector, a producer material (typically lead) which multiplied the number of slow neutrons registered by the detector in order to amplify the signal, a moderator to further slow neutrons, and cylindrical proportional counters utilising BF_3 gas (Simpson, 1948; Simpson et al., 1953). A schematic diagram of the detector is shown in Figure 1.7a.

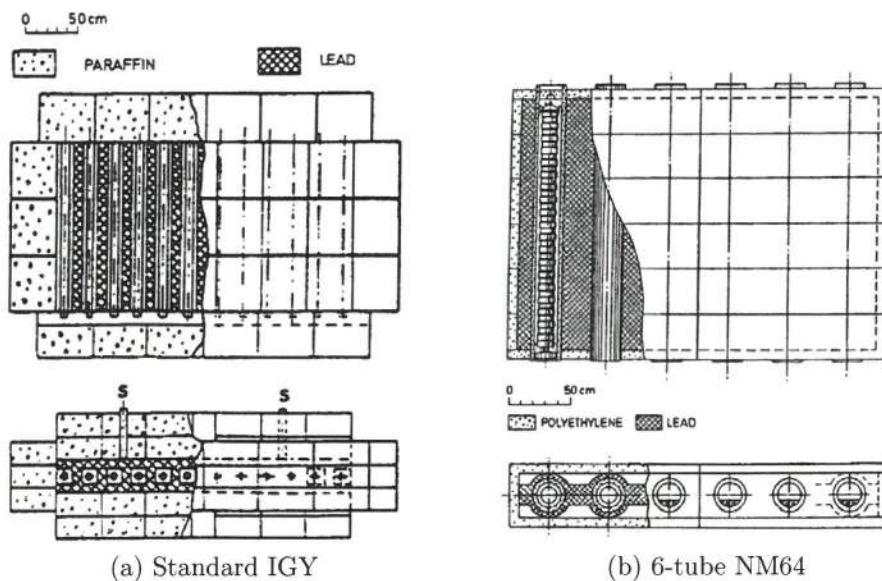


Figure 1.7: Schematic diagrams of two NM configurations: (a) the original Simpson’s 12-tube IGY NM is shown where the paraffin reflector is represented by the outer dotted blocks; the lead producer is represented by the cross-hatched section; the paraffin moderator is represented by the inner dotted blocks; finally the gas-filled proportional counters are denoted by the black circles/tubes. (b) the modern NM64 is shown on the right in its 6-tube configuration where the polyethylene reflector is represented by the outer dotted section; the lead producer is represented by the cross-hatched area; within the producer is a cylindrical polyethylene moderator denoted by the dotted ring; finally in the centre of each tube is a gas-filled proportional counter. In each figure the top schematic is a top-down drawing and the bottom is an end-on drawing. Taken from Kang et al. (2012).

An improved “NM64” NM design is now the preferred detector type, which makes use of a polyethylene reflector, lead producer, polyethylene moderator, and $^{10}\text{BF}_3$ or ^3He gas-filled cylindrical proportional counters (Kang et al., 2012). A schematic diagram of the detector is shown in Figure 1.7b. The new design provided an improvement over the IGY design by a factor of about 3.3 in the count rate per unit area of producer (Stoker et al., 2000); the choice of $^{10}\text{BF}_3$ gas or ^3He gas does not significantly affect the detector performance (Kang et al., 2012).

Muon Detectors

Muon detectors are an example of a scintillation detector, whereby light emitted by atoms excited in a medium is collected and converted into an electrical signal. The scintillation medium can be solid, liquid, or gas; however solid scintillation detectors are attractive as they have a higher electron density (Gloeckler, 2010).

The general configuration of a MD is shown in Figure 1.8. When an energetic particle passes through a scintillator material, some of the particle’s energy is lost in ionising the scintillator material and the scintillator material releases photons. The light pipe directs the photons towards a Photo Multiplier Tube (PMT) where a cascade of electrons are produced and the resulting electrical signal is amplified and recorded through the back-end electronics.

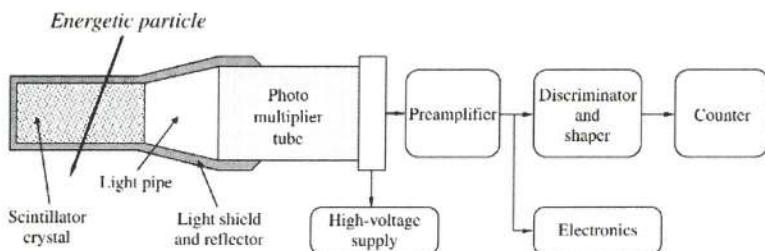


Figure 1.8: Schematic design of a typical scintillation muon detector with back-end electronics (Gloeckler, 2010)

Desirable properties of scintillator materials are a high conversion efficiency, transparency to the light that they emit, short fluorescent decay times, and spectral distributions suitable for photosensitive devices (Gloeckler, 2010).

A range of different material types are used in scintillator detectors however the most common scintillator materials for MDs are organic scintillators consisting of aromatic hydrocarbons (the fluors) in a solid plastic solvent (the base) (Gloeckler, 2010; Fokkema, 2012). Energetic particles traversing the scintillator excite the base rather than the fluor due to the low fluor density. The plastic base however has a low yield and is not transparent to its own scintillation light; thus the fluor is added to therefore increase the yield of this popular type of scintillator (Fokkema, 2012).

1.3.4 Cosmic Ray Observations of Solar Activity and Space Weather

It has long been established that there exists an anti-correlation between GCR intensity and the level of solar activity, over the \sim 11-year period (Forbush, 1958; Parker, 1965; Usoskin et al., 1998; Van Allen, 2000). We also see the 22-year Hale cycle manifesting in the GCR intensity, as interchanging peaked and flat-topped cycles of GCR intensity due to combination of solar activity and CR transport processes (Aslam & Badruddin, 2012; Thomas et al., 2014a), as seen in Figure 1.9. The effects of solar activity on CRs observations are discussed in more details in the introduction to Chapter 4, since the analysis and interpretation of GCR data is presented in detail in that chapter.

There have been many documented observations of space weather events using CR detectors. The most notable types of CR space weather effects are Forbush Decreases (FDs) and Ground Level Enhancements (GLEs); here we discuss the properties of each.

Forbush Decreases

Short-term decreases in the GCR flux were first observed by Forbush (1937) and therefore were later coined as FDs or Forbush Effects (FEs). FDs are characterised by a sharp decrease in GCR intensity over a period of several hours^{to}~~days~~, followed by a gradual recovery taking place over several days^V~~days~~ a week (Cane, 2000; Belov,

for up to

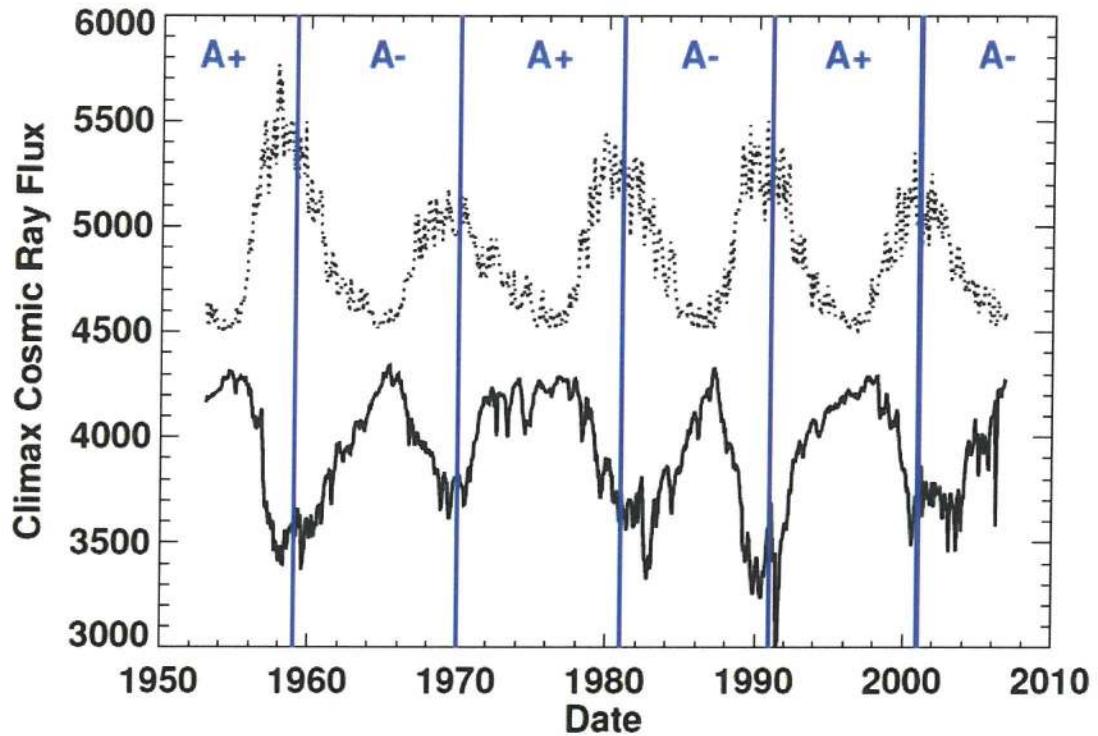


Figure 1.9: Cosmic ray flux measured at the Climax NM (solid line) and rescaled SSN (dotted line) taken from [Hathaway \(2015\)](#). The vertical lines denote changes in the Sun’s global magnetic field polarity, where: A+ indicates positive polarity at the North Pole; A- indicates negative polarity at the North Pole.

2008; [Wawrzynszak & Alania, 2010](#)), as shown in Figure 1.10.

There are two FD origins: one caused by Corotating Interaction Regions (CIRs) ([Dumbović et al., 2016](#)), and one caused by Interplanetary Coronal Mass Ejections (ICMEs) and the shocks they drive ([Belov, 2008](#)). The biggest FDs (magnitudes $> 5\%$) are strictly associated with ICMEs ([Belov et al., 2001](#)). Of the kind caused by ICMEs, the majority are produced by ICMEs with speeds in the range 400–1200 km s $^{-1}$ ([Lingri et al., 2016](#)); the typical speed of the solar wind is, for slow solar wind, in the range 300–400 km s $^{-1}$, and for fast solar wind, ~ 750 km s $^{-1}$ ([Owens & Forsyth, 2013](#)). Previous literature has also shown that the type caused by CIRs produce recurrent, more symmetric, and lower-amplitude decreases ([Dumbović et al., 2012](#)), while the type caused by ICMEs result in the more strongly asymmetric decreases as shown in Figure 1.10 ([Lockwood, 1971](#); [Cane, 2000](#); [Dumbović et al., 2012](#)). In addition the ICME-driven FDs typically result in a two-step FD, where

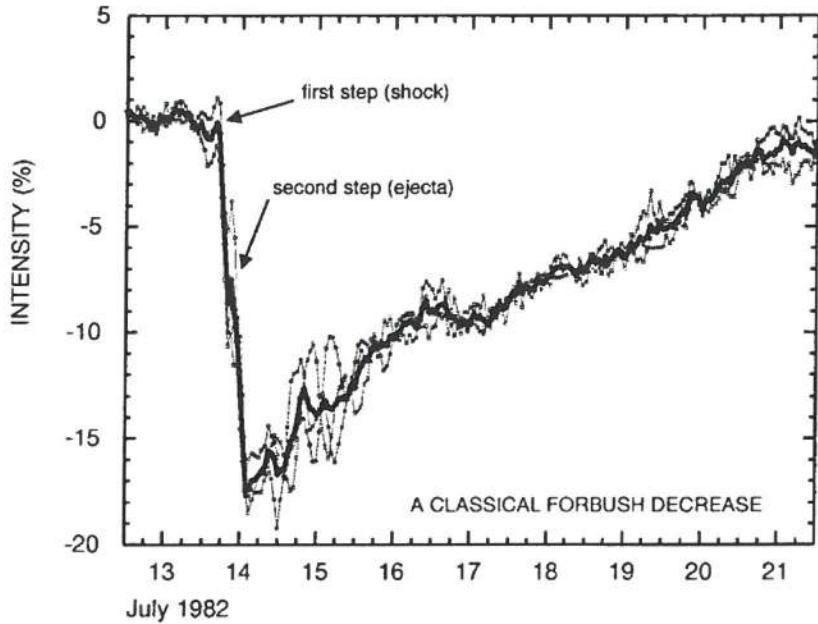


Figure 1.10: A two-step Forbush decrease measured at three NM stations, Deep River, Mt. Wellington, Kerguelen, in July 1982 (Cane, 2000). The thicker black line indicates the average of the count rates from the three stations. Arrows show the start of the two decreases caused by the shock and the ICME ejecta.

the first step of the decrease is due to the passage of the leading shock and the second step is due to the ICME itself, as shown in Figure 1.11 (Cane, 2000).

Lockwood (1971) showed that there is a rigidity dependence on the amplitudes of FDs, which is approximately related to $R^{-\gamma}$, where the exponent ranges from $0.4 \lesssim \gamma \lesssim 1.2$. In addition, Belov et al. (2001, 2014) showed the magnitude of the FD is proportional to the speed, mass, and width of the Coronal Mass Ejection (CME).

The Forbush Effects and Interplanetary-disturbances Database (FEID)² is a record of all the FDs observed since the beginning of the Global Neutron Monitor Network (GNMN) (Mishev & Usoskin, 2020). The total number of events is ~ 7630 during the epoch 1957–2020. Many studies have analysed observations of FDs using these data and investigated their features, driving factors, and precursors; for an overview see: Belov et al. (2001); Usoskin et al. (2008); Wawrzynczak & Alania (2010); Rockenbach et al. (2014); Arunbabu et al. (2015).

²<http://spaceweather.izmiran.ru/eng/dbs.html>

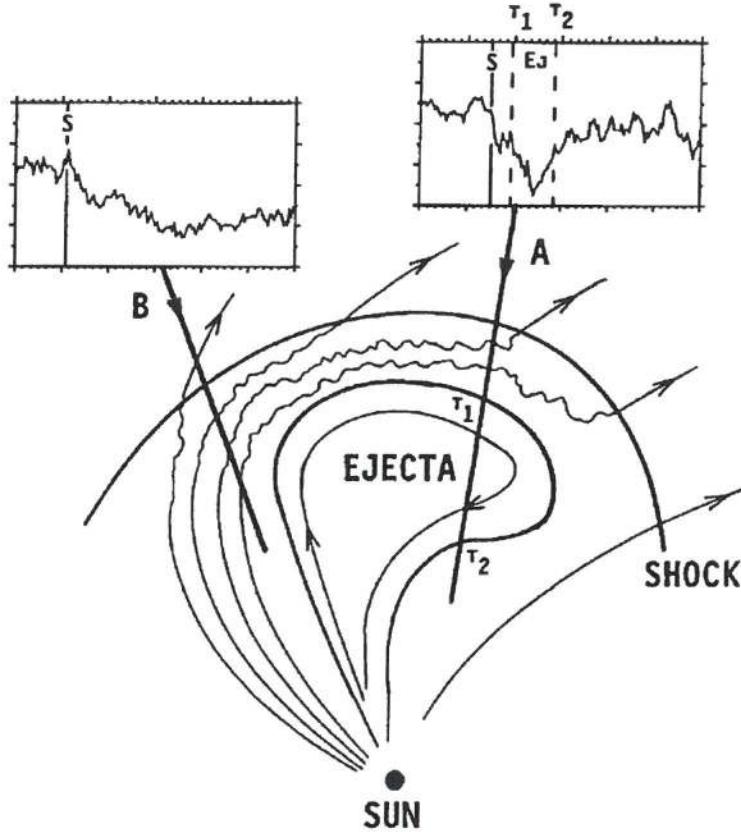


Figure 1.11: A schematic diagram of an ICME-driven FD taken from Cane (2000). It shows the different cosmic ray responses from two paths, indicated by A and B. Path A experiences the shock and ejecta, therefore experiencing a two-step FD; B only experiences the shock, therefore experiencing a single decrease. The time of shock passage is indicated by a solid, vertical line marked, S; the start and end times of ejecta passage are indicated by vertical, dashed lines marked T1 and T2, respectively.

Ground Level Enhancements

Short-term increases in the GCR flux were first observed in the 1940s and early 1950s, but it was not until after the largest recorded event, in September 1956, that these increases were defined as GLEs (Cramp, 1996). GLEs are the detection of an increased number of the highest-energy portion (> 500 MeV, Kuwabara et al., 2006b) of SEPs arriving at Earth along lines of Interplanetary Magnetic Field (IMF) that are accelerated following a solar eruptive event (McCracken et al., 2012; Polianov & Usoskin, 2017). The SEPs, which cause GLEs, can cause serious damage to satellite electronics and are a hazard to air crew and astronauts; hence, the monitoring of these events are of importance for space weather forecasting. GLEs

are characterised by a sharp rise in CR intensity over a period of several minutes^{in to} hours, followed by a gradual decay taking place over several hours, as shown in Figure 1.12.

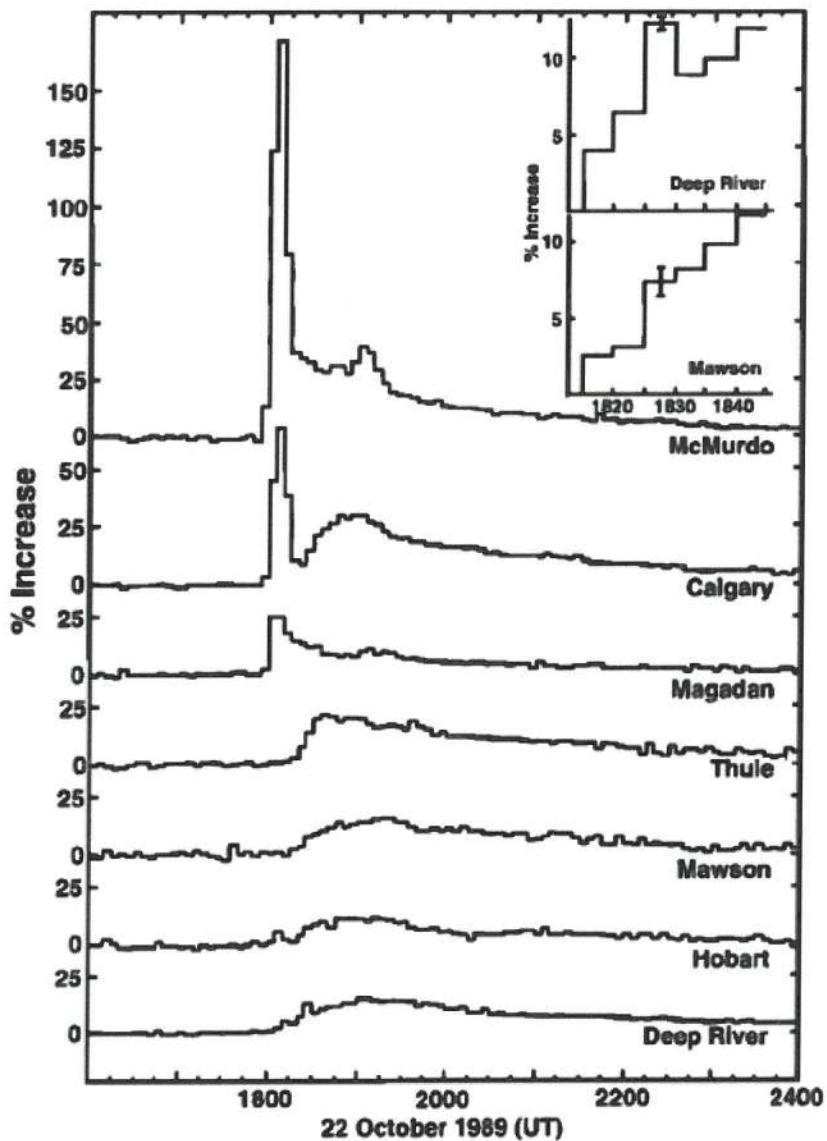


Figure 1.12: A GLE measured at nine NM stations in October 1989, taken from Cramp, J. L. et al. (1997).

GLEs result from energetic solar eruptions such as flares and CMEs (McCracken et al., 2012). The total number of GLEs observed to-date is low: there have been only 72. The GLE database³ is a record of events, starting from GLE 5 (February

³<https://gle.oulu.fi>

1956), since the beginning of the GNMN (Usoskin et al., 2016). Many studies have discussed the observations of GLEs, investigating their features as well as the spectra and anisotropy of PCRs that produce the GLEs; for an overview see: Shea & Smart (1982); Cramp (1996); Belov et al. (2010); McCracken et al. (2012); Strauss et al. (2017); Mishev et al. (2018). Strauss et al. (2017) analysed the shapes of fourteen GLEs and showed the existence of a linear dependence between the rise (τ_r) and decay (τ_d) times which they empirically determined to be $\tau_d \approx 3.5\tau_r$. In addition, the average decay time of GLEs as measured by Strauss et al. (2017) was $1.8^{+1.9}_{-1.3}$ hours.

The solar magnetic field is ‘frozen’ into the solar wind plasma. As the Sun rotates, so do the IMF lines which forms an Archimedean spiral, the Parker spiral (Parker, 1958; Parker & Jokipii, 1976). A curved field line connecting the Sun to the Earth stems from the western limb of the Sun, at a longitude of about 60° , which is known as the ‘garden hose’ field line (see Fig. 1.13) (Duldig et al., 1993; Hathaway, 2015). Charged particles follow magnetic field lines and therefore SEPs that are accelerated in flares located near to the solar end of the ‘garden hose’ field line usually arrive at Earth rapidly and have very sharp onsets (Duldig et al., 1993; Andriopoulou et al., 2011). This causes a strong anisotropy in the arrival directions of the early SEPs inducing GLEs, as shown in Figure 1.12. The McMurdo, Calgary, and Magaden NM stations observed an earlier GLE onset, with a high magnitude, than the other stations (Duldig et al., 1993; Cramp, J. L. et al., 1997). This shows that for a sufficiently anisotropic event, i.e. a GLE where SEPs are predominantly directed towards Earth, different stations at the same cut-off rigidity, but located at different longitudes, may respond differently.

Conversely, GLEs associated with flares far from the ‘garden hose’ field line are usually delayed in their arrival at Earth, due to having to cross magnetic field lines, and have more gradual increases to maximum intensity (Duldig et al., 1993). Very few GLEs have originated far away from the base of the ‘garden hose’ at the solar surface (Duldig et al., 1993; Andriopoulou et al., 2011).

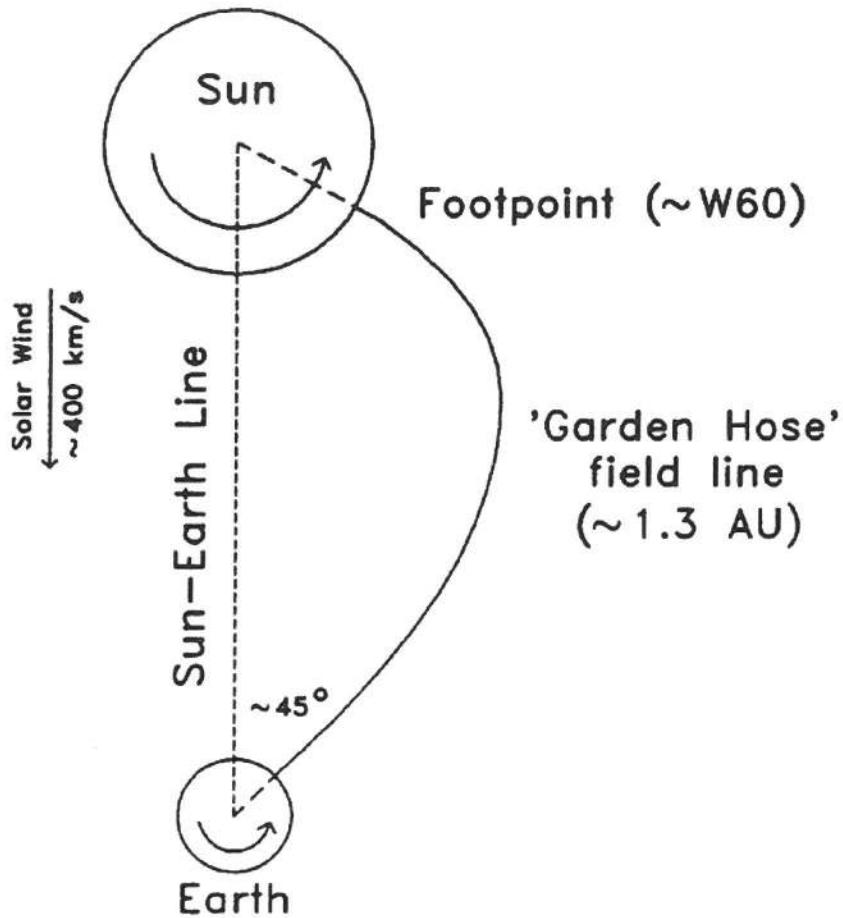


Figure 1.13: A schematic diagram of the ‘garden hose’ field line taken from Duldig et al. (1993).

Most GLEs are detected by multiple stations because of a low enough anisotropy (Strauss et al., 2017; Belov et al., 2018) and in part due to their definition. The accepted definition of a GLE since the 1970s has been (Poluianov et al., 2017):

a GLE event is registered when there are near-time coincident and statistically significant enhancements of the count rates of at least two differentely located NMs.

However, recently a newer GLE definition has been adopted due to the increase in the number of NM stations that are more sensitive to lower energy CRs due to their high latitudes (i.e. in near-polar regions) or higher altitudes. It is a concern that these new NM stations will classify many more GLEs than their near-sea-level counterparts, thus affecting the homogeneity of the current list of GLEs (Poluianov

et al., 2017). Therefore, the new GLE definition is as follows (Poluianov et al., 2017):

a GLE event is registered when there are near-time coincident and statistically significant enhancements of the count rates of at least two differently located neutron monitors, including at least one neutron monitor near sea-level and a corresponding enhancement in the proton flux measured by a space-borne instrument(s).

The new definition also invoked the introduction of a sub-GLE class, defined as (Poluianov et al., 2017):

a sub-GLE event is registered when there are near-time coincident and statistically significant enhancements of the count rates of at least two differently located high-elevation neutron monitors and a corresponding enhancement in the proton flux measured by a space-borne instrument(s), but no statistically significant enhancement in the count rates of neutron monitors near sea level.

Finally, a GLE real-time alarm system was developed by Kuwabara et al. (2006a,b), using data from NMs and MDs, which has been shown to provide the earliest alert for the onset of SEP-driven space weather events. They showed their alerts provide a warning up to an hour earlier than the storm onset. Furthermore, they also show that through utilising the GNMN, monitoring precursory anisotropy, they can also issue warnings several hours ahead of near-Earth, in-situ satellite observations. They state that using both NMs and MDs provides a dual energy range for observations, providing a more effective system.

Please include here some discussion on GLE impacts
e.g. the SWX impacts.
In particular I'm thinking the Hubert ^{et al} 2020 paper about
avionics.

1.4 The HiSPARC Experiment

1.4.1 Background

The High School Project on Astrophysics and Research with Cosmics (HiSPARC) is a scientific outreach project that was initiated in the Netherlands in 2002 (Bartels, 2012). The HiSPARC experiment has two main goals: the study of Ultra-High-Energy Cosmic Ray (UHECR) for astroparticle physics research, and to serve as a resource to expose high school students to scientific research (Bartels, 2012).

HiSPARC is a global network of muon detectors spread across the Netherlands, Denmark, the UK, and Namibia. There are \sim 140 stations in the HiSPARC network (van Dam et al., 2020b) which have been uploading data for varying durations since 2005. The detection philosophy of HiSPARC is to sample the footprints of Extensive Air Showers (EASs) using coincident triggers between scintillation detectors. The detectors at each station record muon counts and may be used for many scientific experiments, such as: reconstruction of the direction of a cosmic ray induced air shower, reconstruction of the energy of the air shower's primary particle, investigation between the atmospheric conditions and the number of cosmic rays observed, etc. A comprehensive review of the HiSPARC experiment is provided by Fokkema (2012) and van Dam et al. (2020b).

The HiSPARC network has predominantly been used to study UHECRs, i.e. Primary Cosmic Rays (PCRs) with energies in excess of $\sim 10^{14}$ eV. However, van Dam et al. (2020a) used the HiSPARC data to derive the Cosmic Ray (CR) flux at sea level for PCRs with energies between 10^{12} ~~and~~ 10^{16} eV. Furthermore, Fan & Velthuis (2018) provided a study into the anti-correlation between atmospheric pressure and the HiSPARC data, as well as claiming to demonstrate the correlation between the daily-average of CR events and solar activity proxies.

? Did they demonstrate it or do you think there is a flaw/issue with the result?

1.4.2 HiSPARC Detector and Station Configuration

As HiSPARC was set up as an outreach programme for high schools, this impacted detector design (Fokkema, 2012; van Dam et al., 2020b). Resources are limited in schools and the detectors are usually financed by the participating high schools, colleges, and universities. In addition, students (accompanied by their teachers and local node support staff) are responsible for assembly and installation of their detectors, which are typically installed on the roofs of schools. Due to this, the detectors needed to be cheap, robust, and easily maintainable, therefore the scintillation detector was selected for the HiSPARC network.

Scintillators consist of materials that emit light when charged particles pass through them with sufficient energy to ionise the scintillator material. The total light produced is proportional to the number of charged particles, and can be collected by a Photo Multiplier Tube (PMT). Each HiSPARC detector utilises a plastic scintillator of dimensions $1000\text{ mm} \times 500\text{ mm} \times 20\text{ mm}$, providing a detection area of 0.5 m^2 . A vertically incident Minimum Ionising Particle (MIP) has a most probable energy loss in 2 cm of the scintillation material of 3.51 MeV ($\equiv 1\text{ MIP}$) (van Dam et al., 2020b).

The scintillator is glued to a triangular/‘fish-tailed’ light-guide (dimensions, base: 500 mm; top: 25 mm; height: 675 mm), and a light-guide adapter provides the optical interface between the square end of the light-guide and the cylindrical aperture of the PMT. The configuration of a single HiSPARC detector is shown in Figure 1.14.

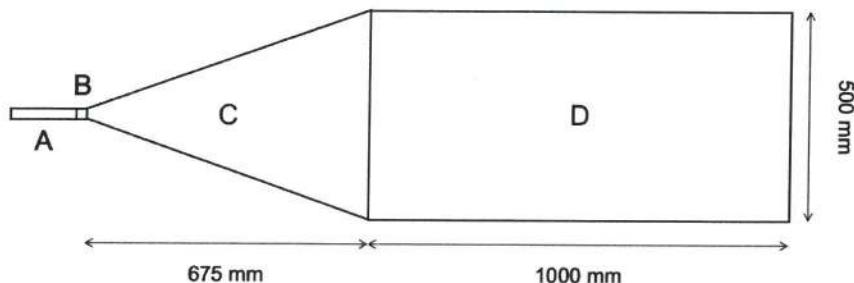


Figure 1.14: Schematic diagram of the HiSPARC scintillation detector. (A): PMT; (B): light-guide adaptor; (C): light-guide; (D): scintillator.

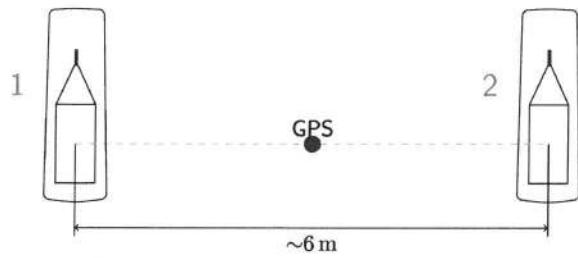
The scintillator is made of a material consisting of polyvinyltoluene as the base, with anthracene as the fluor, and the emission spectrum peaks at a wavelength of 425 nm (Fokkema, 2012; Bartels, 2012). The light-guide is made from Polymethylmethacrylate (PMMA) and has a comparable refractive index to the scintillator (1.58 and 1.49, respectively), reducing refraction effects between the two materials (van Dam et al., 2020b).

The PMT used is an ETEEnterprises 9125B model, with a 25 mm aperture, blue-green sensitive bialkali photocathode, and 11 high-gain dynodes (Bartels, 2012; ETEEnterprises, 2020). The quantum efficiency of the PMT used in the HiSPARC detectors peaks at around 375 nm at 28%, and at 425 nm the quantum efficiency is 25% (Fokkema, 2012).

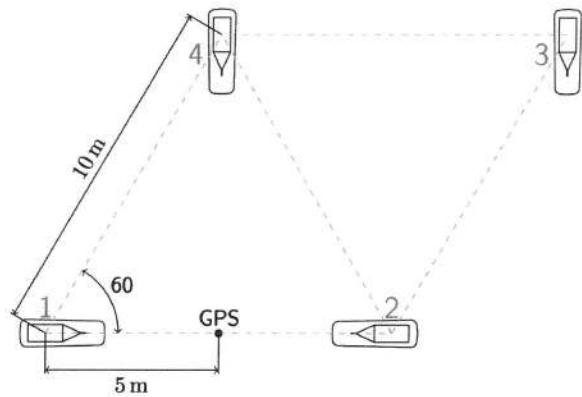
Each detector is wrapped in aluminium foil (thickness 30 µm) and a black, vinyl material (thickness 0.45 mm), which is usually used as a pond liner, to ensure light-tight detectors and to reduce the noise level from stray photons (van Dam et al., 2020b). In addition, each detector is placed inside its own a plastic roof box to again ensure that it is light-tight, and to also ensure that it is weather-proof, as the detectors are usually located on the roofs of schools, colleges, and universities.

The HiSPARC detectors have, individually, a high muon-detection efficiency, close to 100% (Fokkema, 2012; van Dam et al., 2020b), therefore they are capable of observing any muons that traverse them. A HiSPARC station combines either ~~2~~^{two} or ~~4~~^{four} detectors, to observe coincident muons ('events'), and typical configurations of each are shown in Figure 1.15. The separation between detectors varies from station-to-station and this influences the measurable footprint and hence the observable PCRs.

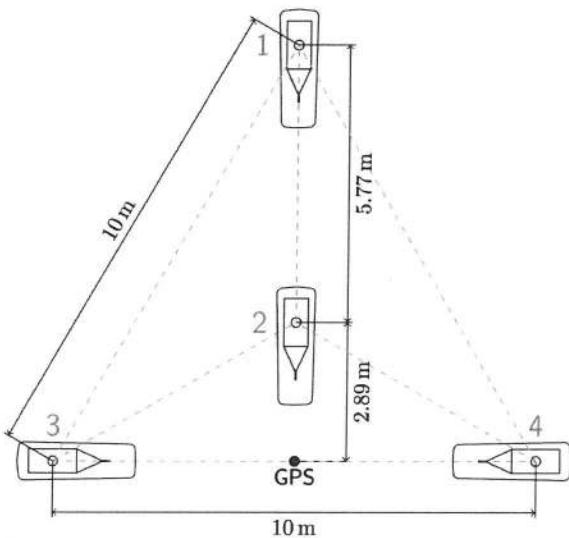
Furthermore some stations have the capability to measure the local atmospheric properties, such as temperature, pressure, relative humidity etc. Moreover, some stations also record the 'singles' rates, i.e. the frequency at which an individual detector is triggered, independently of the other detectors in the station. The singles



(a) Two-detector station configuration



(b) Four-detector station configuration (triangle arrangement)



(c) Four-detector station configuration (diamond arrangement)

Figure 1.15: Typical formations of two-detector and four-detector stations (Fokkema, 2012; van Dam et al., 2020b). In each, the black circle denotes a GPS antenna which is located in between the detectors to provide a precise timestamp for each signal.

rates are important when investigating non-EAS events.

The scientific goals that can be achieved also vary between the two- and four-detector stations. When at least three detectors in a four-detector station observe particles of an EAS, the direction of the EAS (and thus the direction of the PCR) can be acquired using triangulation calculations. When only two detectors in a station observe particles of an EAS it is only possible to reconstruct the arrival direction along the axis that connects the centres of those two detectors (thus it is not possible to reconstruct the direction of the PCR).

The PMTs of the detector in a station are connected to a HiSPARC electronics box, which samples and digitises the signal at a rate of 400 MHz, and each PMTs is connected to the electronics box using cables of a standard length of 30 m, to minimise any timing offsets between detectors (Fokkema, 2012; van Dam et al., 2020b). A schematic diagram showing the configuration of, and interfaces between, the HiSPARC hardware is shown in Figure 1.16. The electronics boxes are capable of controlling and reading two PMTs (see Fig. 1.16), therefore a four-detector station requires two electronics boxes: a primary and a secondary.

Figure 1.16 shows all the hardware interfaces with the HiSPARC electronics box, which communicates with the PC via USB. There are two connections for each PMT, one for the control, i.e. High Voltage (HV), and another for the signal. In addition, a Global Positioning System (GPS) antenna is located in between the detectors in the station (shown in Fig. 1.15). The HiSPARC electronics box contains a GPS board, which provides an accurate timestamp for the data (Fokkema, 2012).

1.4.3 HiSPARC Data Acquisition

HiSPARC Data Acquisition (DAQ) software is used to control and read-out from the HiSPARC electronics box. The DAQ software was developed using LabVIEW, to be executed on a Windows PC (van Dam et al., 2020b). Figure 1.17a shows the typical output signal from one of the PMTs, recorded by the DAQ software.

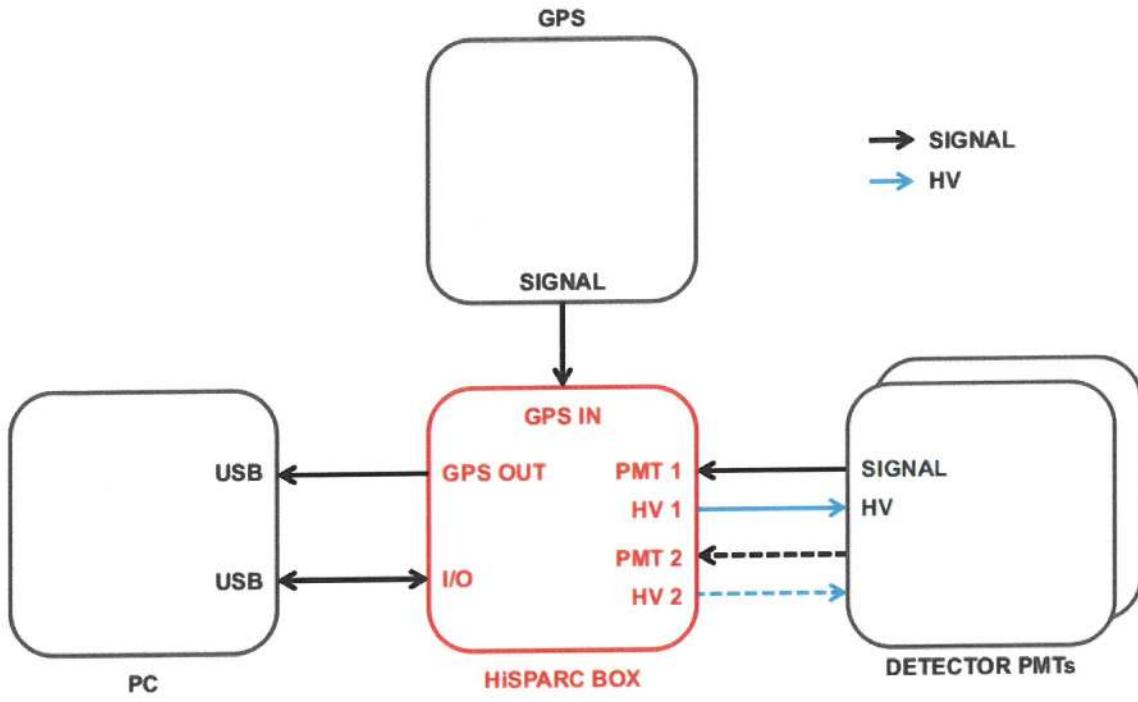


Figure 1.16: A schematic diagram showing the configuration and interfaces between the HiSPARC hardware for a two-detector station.

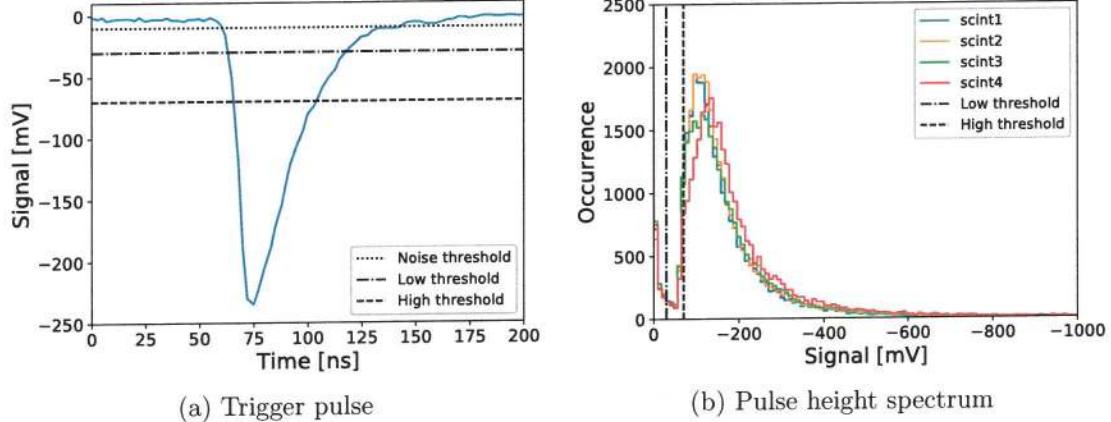


Figure 1.17: (a): An example PMT signal after digital conversion by the HiSPARC electronics box. The horizontal lines denote: the noise cut-off (dotted line), which is used for setting a limit when integrating the pulse height, to give the pulse integral; the low-voltage threshold (dash-dot); the high-voltage threshold (dashed). The role of the high- and low-voltage thresholds are described in the text below. (b) The pulse height distribution over the course of a single day from HiSPARC station 501. The vertical lines show the low-voltage threshold (dash-dot) and the high-voltage threshold (dashed).

The depth of the trace is called the *pulse height* and the area under the curve, the *pulse integral*. The pulse integral is a measure of the number of scintillation photons that have arrived at the PMT, exceeding the noise threshold (-10 mV)

(van Dam et al., 2020b). The HiSPARC DAQ software determines a signal baseline of the PMT (i.e. the background signal without an incident muon), the pulse height, and pulse integral.

The pulse height spectrum (see Fig. 1.17b) is a histogram of all the pulses recorded by a detector. It is composed of two main regions: the left side which falls off rather steeply, and the main, asymmetric part of the spectrum which features a peak and a long tail. The left side of the spectrum is understood to be from high-energy photons (gamma rays) produced in air showers (Fokkema, 2012). These high-energy photons may undergo pair production when interacting with the scintillator which may produce ionising electron and positron pairs.

The main, asymmetric distribution, which features a peak and a tail, is from charged particles (i.e. muons and electrons) (van Dam et al., 2020b). The mean energy loss of particles in a material is described by the Bethe-Bloch formula (Bethe, 1932; Ziegler, 1999); however, this does not account for fluctuations in energy loss and a Landau distribution describes the fluctuations in energy loss of particles (Fokkema, 2012). Due to the resolution of the HiSPARC detectors the distribution in Figure 1.17b is best described by the convolution of the Landau distribution and a normal distribution which describes the resolution of the detector (Fokkema, 2012). The peak of the distribution, the Most Probable Value (MPV), is the most likely energy lost by a particle in the detector, i.e. the 3.51 MeV (van Dam et al., 2020b). It has been shown that the location of the MPV can vary due to the effects of atmospheric temperature (Bartels, 2012; van Dam et al., 2020b).

For each PMT-channel, two discriminator thresholds can be defined: low- and high-voltage, as shown in Figure 1.17b, as vertical dash-dot and dashed lines, respectively. The trigger thresholds are placed to reject the noise signals from the data, i.e. the left side of the pulse height spectrum. If a signal exceeds the high threshold, there is a high probability that the signal was generated by a particle in the detector. The ‘singles’ data counts any time that the signal measured from a

PMT exceeds the low- or high-threshold.

The HiSPARC experiment is configured in such a way as to ensure that each station across the HiSPARC network measures a similar count rate of muons, in order to aid the direct comparison between the different stations in the network. When configuring the station, a trigger threshold must be applied for the PMT signals. This is standardised across the HiSPARC network and can be seen in relation to a detector trigger pulse in Figure 1.17a. There are two thresholds, low: -30 mV, which represents ~ 0.2 of a MIP; high: -70 mV, which represents ~ 0.5 of a MIP (Fokkema, 2012; van Dam et al., 2020b). van Dam et al. (2020b) states the thresholds were chosen to increase the sensitivity for observing gamma rays and low-energy electrons; the sensitivity is lower than for muons. However, as the detectors can measure gamma rays and electrons, it can be difficult to determine whether an individual detection was a muon, or another MIP, which is why the HiSPARC network usually relies on detecting ‘events’, from coincident muons.

To register an ‘event’, the detectors must measure signals which exceed the low threshold, within a limited time window. The time-line of the data acquisition during an event is given in Figure 1.18.

When channel 1 (i.e. PMT 1) exceeds the low-threshold a coincidence window is opened. The length of the coincidence window is ~ 1.5 μ s. If, within this window, channel 2 (i.e. PMT 2) also measures a signal that exceeds the low threshold, the trigger condition is met and an event is generated. When a trigger is issued, the output is stored in the buffer of the Field Programmable Gate Array (FPGA) in the HiSPARC electronics box (Fokkema, 2012). The data acquisition software reads the data from the buffer and the full event data consist of: (i) data measured before the coincidence window was opened (the pre-trigger window, typical length ~ 1 μ s); (ii) data measured within the coincidence window; (iii) data measured after the trigger period (the post-trigger window, typical length ~ 3.5 μ s). Therefore, the event window lasts for ~ 6 μ s in total, with the maximum length of an event window of

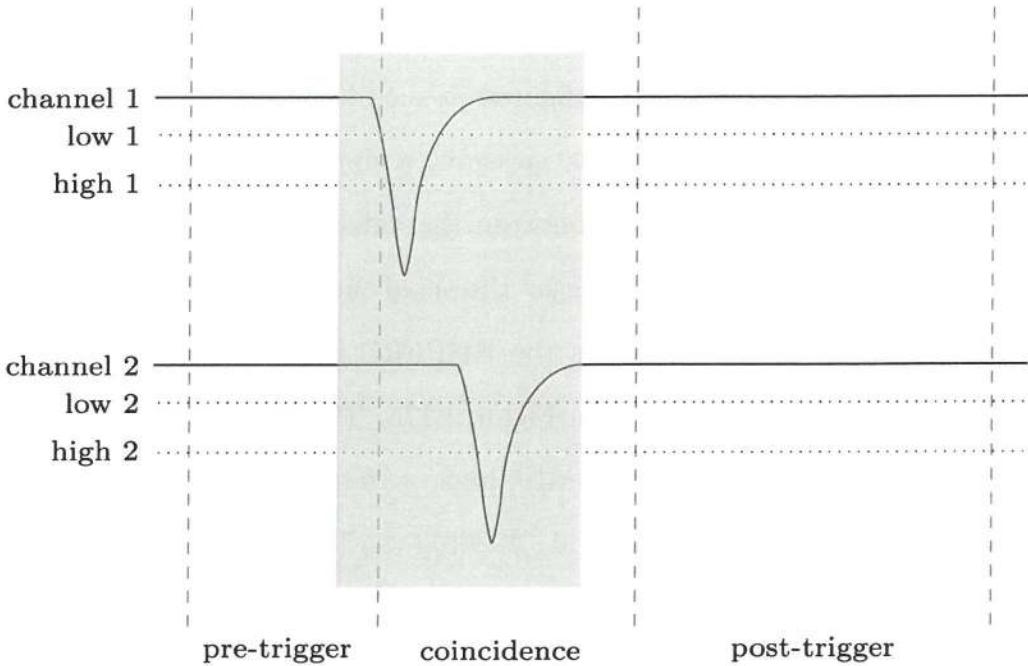


Figure 1.18: Schematic data acquisition of an event (Fokkema, 2012). The dashed vertical lines denote the epochs of the pre-trigger, coincidence, and post-trigger windows. The grey, shaded region shows the data reduction window and data outside this window are not stored. The dotted, horizontal lines denote the low- and high-voltage thresholds.

10 μs (van Dam et al., 2020b). Finally, a data reduction algorithm is applied on the full event window which determines the part of the signal containing the muon pulses—compared to the baseline—and removes the rest, thus greatly reducing the size of the event data (Fokkema, 2012).

The default trigger condition for detecting an air shower event between multiple PMTs within a station differs for a two- and four-detector station. In a two-detector station, an event is recorded if the PMT signals from both detectors exceed the low threshold within the coincidence time window (1.5 μs). In a four-detector station, the default trigger condition is either: (i) at least two detectors exceed the high threshold within the coincidence time window; (ii) at least three detectors exceed the low threshold within the coincidence time window. These are the default conditions, but there are other, user configurable ways of triggering the station. Under these nominal conditions, each HiSPARC station triggers events at a rate of ~ 1 Hz, i.e. substantially lower than the muon flux stated in Section 1.3.2.

Data recorded by the HiSPARC stations are stored and are available on the HiSPARC Public Database⁴, where each station is listed, grouped by local nodes. For every station one can see its ID, name, and a coloured square and circle displaying its current data delivery and DAQ status, respectively. Clicking on any station takes you to a dedicated page which displays its data on a user-selected day. Where data are available, it is possible to download:

- events rate data: where multiple detectors in a station are triggered to satisfy that station's trigger condition;
- singles rate data: the count rates of the individual detectors within a station;
- weather data: meteorological data, including pressure and temperature;
- coincidences data: the counts where different stations measure the same event (to within 1.36 μ s); it is possible to determine if stations measured the same event by comparing the GPS timestamps of events.

To support the HiSPARC project, the Simulation and Analysis Program Package for HiSPARC Research and Education (SAPPHiRE) Python package (Fokkema, 2012; Fokkema et al., 2012) was written. This Python package provide a framework to analyse the HiSPARC data, but also an alternative way to acquire the data.

1.5 Thesis Structure

In this thesis, a number of projects are presented which explore the themes of understanding the solar interior-atmosphere linkage and space weather applications. This is broken down into three major projects: a feasibility study on Cosmic Ray (CR) space weather applications, an investigation into the effects of solar activity on CR observations, and a study of the Solar Mean Magnetic Field (SMMF). The thesis is structured as follows:

⁴<https://data.hisparc.nl/>

In Chapter 2 we perform a feasibility study to determine whether the High School Project on Astrophysics and Research with Cosmics (HiSPARC) network is suitable for monitoring space weather events. This was achieved using historic data from the HiSPARC network to search for the signature of space weather events. In addition we performed simulations of Air Showers (ASs) to determine the expected variation in muon counts during space weather events.

Following the results of Chapter 2, Chapter 3 outlines the design and implementation on the University of Birmingham campus of an alternative HiSPARC station, with a novel arrangement of detectors that removes thermally induced, diurnal variations in the data, and reduces the existing energy bias of observable Primary Cosmic Rays (PCRs) using the HiSPARC network. In this chapter we discuss the set-up of the station, review the data and its noise properties, and finally also perform simulations using artificial data to investigate the capabilities of this new station configuration.

In Chapter 4 we studied long-term variations of Galactic Cosmic Ray (GCR) intensity in relation to the Sun Spot Number (SSN) during the most recent solar cycles. This study, which was published in the journal Solar Physics (Ross & Chaplin, 2019), analysed the time lag between the GCR intensity and the SSN, and the hysteresis effect of the GCR count rate against SSN for Solar Cycles 20–24.

Chapter 5 presents a frequency-domain analysis of over 20 years of high-cadence Birmingham Solar Oscillations Network (BiSON) observations of the SMMF. We modelled the power spectrum of the BiSON SMMF data to draw conclusions about the morphology of the SMMF, particularly focusing on the source of the rotationally modulated component in the signal. A significant portion of the work presented in this chapter was published in the journal Monthly Notices of the Royal Astronomical Society (Ross et al., 2021).

In Chapter 6 we further investigated the BiSON SMMF data. Here we examined the residual spectrum, after removing our best-fitting model, to search for evidence

of a magnetic signature of global Rossby modes (r modes).

We finish with concluding remarks in Chapter 7.

All of the work and results presented in this thesis are my own and all the data analysis was performed by me. Input from others came in the form of advice and consultation, in addition to the supply of raw data.

2 HiSPARC as a Space Weather Detector

2.1 Introduction

The observation of Cosmic Rays (CRs) provides a tool to monitor the effects of space weather on Earth. Space weather events have been regularly monitored by ground-based CR detectors since the early 20th Century (Forbush, 1937; Kudela et al., 2000; Schwenn, 2006) and the detectors have been instrumental in characterising the conditions surrounding space weather events.

Short-term increases in the Galactic Cosmic Ray (GCR) flux were first observed in the 1940s and early 1950s, but it was not until after the largest recorded event in September 1956 that these increases were defined as Ground Level Enhancements (GLEs) (Cramp, 1996). GLEs are the detection of an increased number of the highest-energy portion (> 500 MeV, Kuwabara et al. 2006b) of Solar Energetic Particles (SEPs) arriving at Earth following a solar eruptive event (McCracken et al., 2012; Poltovets & Usoskin, 2017). The SEPs, which cause GLEs, can cause serious damage to satellite electronics and are a hazard to air crew and astronauts; hence, the monitoring of these events is of importance for space weather forecasting.

The total number of GLEs observed to-date is low: there have been only 72. The GLE database¹ is a record of events measured using the Global Neutron Monitor Network (GNMN), starting from GLE 5 (February 1956), since the beginning of CR space weather monitoring operations (Usoskin et al., 2016). Many studies have investigated the observations of GLEs, analysing their characteristics, as well as the

¹<https://gle.oulu.fi>

2 HiSPARC as a Space Weather Detector

2.1 Introduction

The observation of Cosmic Rays (CRs) provides a tool to monitor the effects of space weather on Earth. Space weather events have been regularly monitored by ground-based CR detectors since the early 20th Century (Forbush, 1937; Kudela et al., 2000; Schwenn, 2006) and the detectors have been instrumental in characterising the conditions surrounding space weather events.

largest GCR
flux?

Short-term increases in the Galactic Cosmic Ray (GCR) flux were first observed in the 1940s and early 1950s, but it was not until after the largest recorded event in September 1956 that these increases were defined as Ground Level Enhancements (GLEs) (Cramp, 1996). GLEs are the detection of an increased number of the highest-energy portion (> 500 MeV, Kuwabara et al. 2006b) of Solar Energetic Particles (SEPs) arriving at Earth following a solar eruptive event (McCracken et al., 2012; Polianov & Usoskin, 2017). The SEPs, which cause GLEs, can cause serious damage to satellite electronics and are a hazard to air crew and astronauts; hence, the monitoring of these events is of importance for space weather forecasting.

The total number of GLEs observed to-date is low: there have been only 72. The GLE database¹ is a record of events measured using the Global Neutron Monitor Network (GNMN), starting from GLE 5 (February 1956), since the beginning of CR space weather monitoring operations (Usoskin et al., 2016). Many studies have investigated the observations of GLEs, analysing their characteristics, as well as the

¹<https://gle.oulu.fi>

of a magnetic signature of global Rossby modes (*r* modes).

We finish with concluding remarks in Chapter 7.

All of the work and results presented in this thesis are my own and all the data analysis was performed by me. Input from others came in the form of advice and consultation, in addition to the supply of raw data.

spectra and anisotropy of Primary Cosmic Rays (PCRs) that produce the GLEs; for an overview see: [Shea & Smart \(1982\)](#); [Cramp \(1996\)](#); [Belov et al. \(2010\)](#); [McCracken et al. \(2012\)](#); [Strauss et al. \(2017\)](#); [Mishev et al. \(2018\)](#).

In addition, a GLE real-time alarm system was developed by [Kuwabara et al. \(2006a,b\)](#), using data from Neutron Monitors (NMs) and Muon Detectors (MDs), which has been shown to provide the earliest alert for the onset of SEP-driven space weather events. GLEs are capable of triggering multiple ground-based detectors at the same time and they showed their alerts provide a warning up to an hour earlier than the storm onset. Furthermore, they also show that through utilising the GNMN, monitoring precursory anisotropy, they can also issue warnings several hours ahead of near-Earth, in-situ satellite observations. They state that using both NMs and MDs provides a dual energy range for observations, providing a more effective system.

On this dual energy range for space weather observations; NMs generally observe PCRs with energies \sim 1–10 GeV and above, while MDs typically observe higher energy PCRs with energies on the order of \gtrsim 10 GeV ([Kuwabara et al., 2006a](#); [Rockenbach et al., 2014](#)). After a solar eruptive event, it is expected that the first particles to arrive at Earth are those with higher energies, i.e. measured by muon detectors ([Lovell et al., 1998](#)), and those that traverse the shortest distance, i.e. travelling along the Interplanetary Magnetic Field (IMF) ([Kuwabara et al., 2006a](#)). Different behaviours observed between MDs and NMs are linked with the different particle species observed; MDs can therefore be linked with the maximum flare energy release and an earlier arrival of the SEPs at Earth ([Kuwabara et al., 2006a](#)).

Short-term decreases in the GCR flux were first observed by [Forbush \(1937\)](#) and therefore were later coined as Forbush Decreases (FDs) or Forbush Effects (FEs). There are two types of FD: one caused by Corotating Interaction Regions (CIRs) ([Dumbović et al., 2016](#)), and one caused by Interplanetary Coronal Mass Ejections (ICMEs) and the shocks they drive ([Belov, 2008](#)). The biggest FDs (magnitudes

> 5%) are strictly associated with ICMEs (Belov et al., 2001). Of the kind caused by ICMEs, the majority of are produced by ICMEs with speeds in the range 400–1200 km s⁻¹ (Lingri et al., 2016); the typical speed of the solar wind is, for slow solar wind, in the range: 300–400 km s⁻¹, and for fast solar wind, \sim 750 km s⁻¹ (Owens & Forsyth, 2013). In addition, Belov et al. (2001, 2014) showed the magnitude of the FD is proportional to the speed, mass, and width of the Coronal Mass Ejection (CME). We also see from the Neutron Monitor Data Base (NMDB) data, generally, the magnitude of the FD is also inversely proportional to the PCR energy.

The Forbush Effects and Interplanetary-disturbances Database (FEID)² is a record of all the FDs observed since the beginning of the GNMN (Belov, 2008). The total number of events is \sim 7630 during the epoch 1957–2020. Many studies have discussed the observations of FDs and investigate their features, driving factors, and precursors; for an overview see: Belov et al. (2001); Usoskin et al. (2008); Wawrzynczak & Alania (2010); Rockenbach et al. (2014); Arunbabu et al. (2015).

The variation in CR counts during space weather events, as measured by several stations with different rigidities across the GNMN, shows a larger GLE or FD magnitude is generally observed for lower rigidity PCRs. Belov et al. (2005) reported the relationship between the variation and the rigidity was approximately $\propto R^{-0.8}$, showing a clear inverse relationship between rigidity and degree of count variation.

Despite most observations of space weather events in the literature utilising data acquired by NMs, there are some reports of observations with MDs also. However, in general, the MD observations of space weather events are significantly less pronounced than the measurements using NMs for both GLEs (Timashkov et al., 2008; Augusto et al., 2016) and FDs (Braun et al., 2009; Rockenbach et al., 2014), due to the higher rigidity PCRs they observe. This suggests that MDs may not be the most suitable instrument for monitoring space weather events. However, increases in muon count rates have been observed of several tens of per cent, as discussed

²<http://spaceweather.izmiran.ru/eng/dbs.html>

by Shea (1990); Lovell et al. (1998); Moraal & Caballero-Lopez (2014). These observations also showed an earlier peak in the MD data than the NM data of up to ~ 1 hour (Lovell et al., 1998; Moraal & Caballero-Lopez, 2014).

The High School Project on Astrophysics and Research with Cosmics (HiSPARC) experiment was set up with the detection philosophy of observing Extensive Air Showers (EASs) of muons, which are typically associated with PCRs with energy of $\sim 10^{14}$ eV and above, that produce large footprints observable with many HiSPARC stations simultaneously (Fokkema, 2012; Bartels, 2012; van Dam et al., 2020b). For PCRs with energies below $\sim 10^{14}$ eV, the induced air shower is small, with almost no observable muon footprint, and for PCRs with energy below $\sim 10^{11}$ eV, there are typically fewer than one or two muons that reach the ground, making their observation difficult (van Dam et al., 2020b). Most muons produced by such low-energy PCRs decay higher in the atmosphere and their energy is mostly transferred into the resultant electron (van Dam et al., 2020b), but depending on the electron energies, they are also observable by HiSPARC as a Minimum Ionising Particle (MIP).

The HiSPARC detectors have, individually, a high muon-detection efficiency close to 100% (van Dam et al., 2020b), therefore they are capable of observing any muons that traverse them. This project was hence motivated by the existing network of MDs which may have the capability of observing the CRs associated with space weather events.

In the literature there exists no previous work which investigates the use of the HiSPARC network to monitor space weather. Previous studies using HiSPARC data have only considered PCRs with energies $> 10^{12}$ eV, therefore several orders of magnitude larger than the energies usually associated with space weather events (Bartels, 2012; van Dam et al., 2020a). This is due to the heritage of the HiSPARC network, as it was set-up to observe Ultra-High-Energy Cosmic Rays (UHECRs). In this work, we provide a feasibility study to investigate whether the existing HiS-

PARC network is capable of observing space weather events.

Few space weather events have been observed over the lifetime of the HiSPARC network; however, Table 2.1 outlines the specific GLEs and FDs that occurred since the beginning of HiSPARC that were investigated in this work. The table also shows, for reference, the magnitude of the CR count variation observed by two NM stations: Oulu, Finland ($R_C=0.81$ GV), and Kiel, Germany ($R_C=2.36$ GV), where R_C is the rigidity cut-off and is the minimum rigidity PCR observable by the stations due to the Earth's geomagnetic field.

Table 2.1: Largest space weather events since the beginning of HiSPARC, which were searched for within the HiSPARC data. The percentage-change columns provide a reference of how much the CR counts observed by the NM stations at Oulu ($R_C=0.81$ GV) and Kiel ($R_C=2.36$ GV) increased by or decreased by, due to the space weather event. More precise times for the event onset can be found within: NMDB (2018) (for GLEs) and Lingri et al. (2016) (for FDs).

GLE Onset	GLE	% Change		FD Onset	% Change	
		Oulu	Kiel		Oulu	Kiel
13/12/2006	70	~ 90%	~ 30%	08/03/2012	~ 10%	~ 10%
17/05/2012	71	~ 15%	~ 3%	12/03/2012	~ 3 – 5%	~ 3 – 5%
10/09/2017	72	~ 5%	N/A	14/07/2012	~ 3 – 5%	~ 5 – 10%
				21/12/2014	~ 5 – 10%	~ 5 – 10%
				06/09/2017	~ 1 – 2%	N/A
				07/09/2017	~ 6%	N/A

A number of these are very well known (and well studied) Spwx events. Some very brief additional context on the "cause" of the event would be nice.

The specific events in Table 2.1 were selected as: (i) for the GLEs, they are the only three that occurred in the HiSPARC operational period; (ii) for the FDs, they are among the most recent FDs that result in a NM count-rate variation in excess of $\sim 5\%$ and the largest FDs are likely to be the most promising candidates for observation with HiSPARC. In addition, Timashkov et al. (2008) state that an increase of $0.61 \pm 0.09\%$ was measured with a MD in the Moscow Engineering Physics Institute for GLE 70, where the detector had cut-off rigidity of ~ 2.43 GV.

For reference, and later comparison with the HiSPARC results, in Figure 2.1 we show the GLEs, as observed by the Oulu NM station, in Finland, using data taken from the NMDB (NMDB, 2018). We see in Figure 2.1 that the relative increase

of the CR counts during the GLE was large for GLEs 70 and 71, but much lower during GLE 72. For each GLE, the increase is easily observable by-eye in the data. We expect that if we are to observe any of the GLEs, we shall have the best chance of observing GLE 70. Similarly, we show a plot of the FDs, as observed by the Oulu NM station, in Figure 2.2.

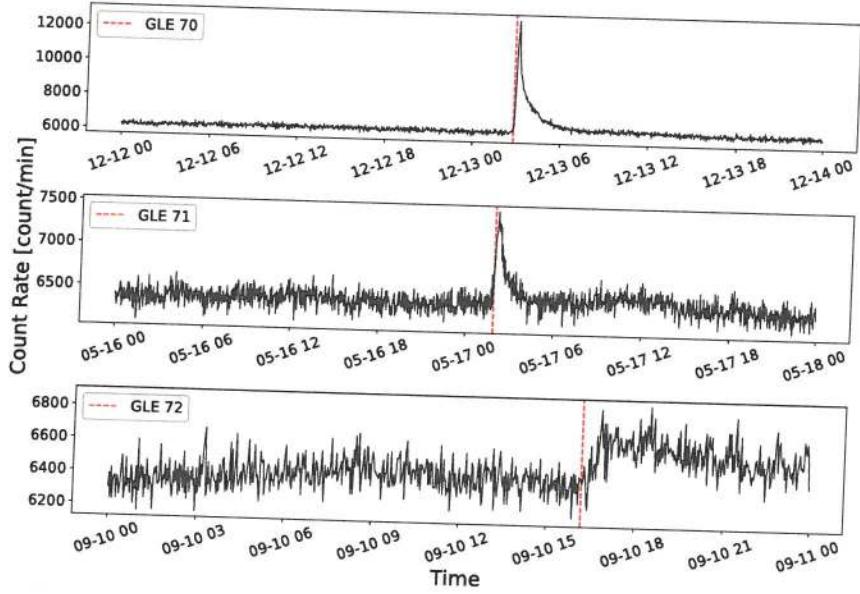


Figure 2.1: GLEs observed by the NM stations based at Oulu. Top panel: GLE 70; middle panel: GLE 71, bottom panel: GLE 72. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-red lines show the epochs of each GLE onset. The units of time on the x-axis are, MM-DD HH.

Each FD in Figure 2.2 produces a moderate decrease in the NM count rate. The relative decreases in the CR counts during the FDs were generally around $\sim 5\%$, and are easily observable by-eye in the data.

2.2 Aims

The principle aim of the project was to determine whether the existing HiSPARC network is capable of observing space weather events. To do this, we initially explored the properties of the HiSPARC detectors in detail, to understand the typical PCRs they observe. In addition, we investigated the data during periods of space

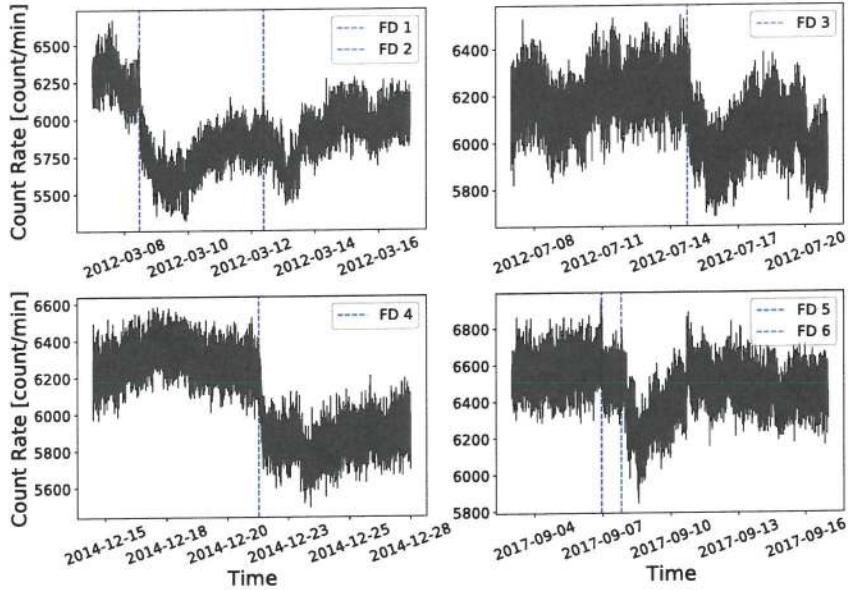


Figure 2.2: FDs observed by the NM stations based at Oulu. Top left panel: FDs during March 2012; top right panel: FD during July 2012; bottom left panel: FD during December 2014; bottom right panel: FD during September 2017. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-blue lines show the epochs of each FD onset. The units of time on the x-axis are, YYYY-MM-DD.

weather activity to search for the associated signatures detailed in Table 2.1. We searched through some of the most reliable HiSPARC stations to determine whether these events were observed in the data. This was done to determine whether, without much effort, we could get a binary answer on whether these events were observed by HiSPARC.

Ground-based observations of muons from air showers are susceptible to the conditions in the atmosphere; where possible, we aimed to correct for these atmospheric effects and again reviewed the corrected data to determine whether the space weather events were observed.

Finally, we also aimed to perform simulations of air showers initiated by CRs to understand the expected muon flux and dispersion at ground level. This would help us to understand how likely it is to observe the PCRs associated with space weather with the HiSPARC detectors, observing muons.

It was highlighted during private communication with the UK Met Office that observations of GLEs are of more interest and importance to space weather fore-

casts and nowcasts. As discussed above, using MDs generally provides observations of higher energy PCRs which relates to the earlier onset of SEPs; thus using the HiSPARC network is of significant interest in this domain. FDs are of lower interest and importance; however, we still searched for FDs within the HiSPARC data for completeness.

2.3 HiSPARC Data

The HiSPARC cosmic ray data are available on the HiSPARC Public Database³, where each station is listed, grouped by local nodes. For every station one can see its ID, name, and a coloured square and circle displaying its current data delivery and data acquisition status, respectively. Clicking on any station takes you to a dedicated page which displays its data on a user-selected day. Where data are available, it is possible to download:

- events rate data: where multiple detectors in a station are triggered to satisfy that station's trigger condition;
- singles rate data: the count rates of the individual detectors within a station;
- weather data: meteorological data, including pressure and temperature;
- coincidences data: the counts where different stations measure the same event (to within 1.36 μ s); it is possible to determine if stations measured the same event by comparing the Global Positioning System (GPS) timestamps of events.

This method of obtaining HiSPARC data is acceptable if only a small quantity of data are needed, but it is cumbersome if large quantities of data are required. To obtain large quantities of HiSPARC data, it was more efficient to use the Event Summary Data (ESD) module within the Simulation and Analysis Program Package for HiSPARC Research and Education (SAPPHiRE) Python package (Fokkema

³<https://data.hisparc.nl/>

et al., 2012). The data are downloaded in the raw HDF5 format, and can then be manipulated using further Python scripts.

A Python script was written, which used the SAPPHiRE ESD module, to request the download of a specific type of data (i.e. events, singles, weather), from a user-specified station, download and open the HDF5 table, manipulate the data to either keep them in the raw cadence or resample into other timebases, and finally store the data in .csv format. This reduced the complexity involved in downloading the data and provided a repeatable method of acquiring the HiSPARC data in a consistent format.

There are \sim 140 stations in the HiSPARC network (van Dam et al., 2020b) which have been uploading data for varying durations since 2005. It was too challenging to acquire and analyse data from every station, hence a smaller sample of 5 stations was selected for investigation. The stations in the sample are outlined in Table 2.2, a mixture of 2-detector and 4-detector stations. Approximately 110 of the \sim 140 stations record singles rates (see Section 1.4), which have only been available since 2016, and only 29 stations acquire meteorological data. In general, throughout the history of the HiSPARC network, the availability of weather data is irregular and many stations that acquire the data go through periods of acquiring no meteorological data at all, which made the selection of stations non-trivial.

The 5 stations in the sample were selected as they generally have both the singles and weather data available, with the exception of station 14001 (University of Birmingham). Station 14001 only came online in 2014; it does not acquire weather data and did not begin acquiring singles data until February 2019, but it was deemed advantageous to include this station as it is maintained by the University of Birmingham, therefore we have full control over the operation of the station and it is a useful reference. Station 501 (Nikhef) is the original station in the network, and serves as the ‘gold standard’ for HiSPARC, therefore it was included. The other three stations all showed good data quality in terms of data availability and consistent operating

? why

conditions. The stations are shown on a map in Figure 2.3.



Figure 2.3: The geographic location of each HiSPARC station considered in this work. Each green circle denotes the location of a detector station.

Figure 2.4 shows the availability of CR data from the sample of stations for each of the space weather events investigated, where purple grids denote no available data, teal denotes events data were available, and yellow denotes events and singles data were available. For each event, we have data available from at least two stations, which allows us to compare the signals. It also shows as the HiSPARC network matures, so does the number of stations with available data, and the type of data available (both events and singles rates).

2.4 Station Properties

2.4.1 Cut-Off Rigidity

To understand the PCR spectrum that ground-based CR detectors may observe, PCR transport simulations are typically performed (McCracken et al., 1962; Plainaki et al., 2009; Danilova et al., 2019). The geomagnetic field can prohibit CR particles from penetrating the magnetosphere and reaching the atmosphere, depending on the particle's energy. As a consequence, the cut-off rigidity is an estimate of the lower



Figure 2.4: The availability of data for each HiSPARC station considered, for each of the space weather epochs listed in Table 2.1. The purple grids denote no available data, teal denotes that only the events data are available, and yellow denotes that both the events and singles data are available.

rigidity-threshold, below which the particle flux is zero due to geomagnetic shielding (measured in Volts, V, or Gigavolts, GV) (Shea et al., 1965; Danilova et al., 2019). Transport simulations allow us to determine the range of PCRs that have sufficient energy to penetrate the Earth's magnetosphere, which reach the atmosphere and produce secondary particle air showers. The PCR spectrum depends strongly on the geographic location of the station; the minimum allowed particle rigidity varies from ~ 17 GV near the equator and theoretically 0 GV at the poles, and the geomagnetic conditions have a strong impact on the PCR spectrum (Shea et al., 1965; Cramp, 1996; Desorgher et al., 2006; Danilova et al., 2019).

Transport simulations typically run a reverse evolution of particles, using a backwards-tracing routine, whereby the particles are simulated from Earth out to the magnetosphere, to determine whether they leave the magnetosphere or remain trapped due to the geomagnetic field (Shea et al., 1965). In this work, to perform the PCR transport simulations we used the PLANETOCOSMICS software (Desorgher, 2005). PLANETOCOSMICS performs Geant4 Monte Carlo simulations of charged particle transport through Earth's magnetosphere based on Störmers trans-

port equation for charged particles (Desorgher, 2005; Desorgher et al., 2006).

PLANETOCOSMICS simulates backward trajectories of charged particles from a given location (latitude, longitude, and altitude) out to the magnetopause for a set of PCR rigidities. For each simulated trajectory there are two possible outcomes: (i) the particles trace out to the magnetopause where they escape Earth's magnetosphere, an allowed trajectory; (ii) the particles are sufficiently bent by the effect of the Earth's magnetosphere that they do not reach the magnetopause and cannot escape the Earth's magnetosphere, a forbidden trajectory (Shea et al., 1965; Desorgher, 2005; Desorgher et al., 2006). The coordinates of the asymptotic direction at the magnetosphere are provided as an output from the simulations. This is the direction of motion of particles upon leaving the magnetosphere, if subjected to no other forces (Shea et al., 1965; Desorgher et al., 2006; Danilova et al., 2019). In this work PLANETOCOSMICS was configured with the Tsyganenko-89 model for the external magnetospheric magnetic field (Tsyganenko, 1989, 2013) and the International Geomagnetic Reference Field (IGRF) internal field model (Thébault et al., 2015).

Each simulated rigidity, whether it followed an allowed or forbidden trajectory, was stored and was used to provide an insight into the rigidity spectrum for a given station. From the allowed trajectories the effective cut-off rigidity (R_C) for the stations was computed, which represents the lower rigidity limit above which cosmic rays can cross the magnetosphere and reach the atmosphere:

$$R_C = R_U - \sum_{i=R_L}^{R_U} \Delta R_i, \quad (2.1)$$

where R_U is the upper rigidity (the last allowed trajectory before the first forbidden trajectory); R_L is the lower rigidity (the last allowed trajectory before which all other trajectories with a lower rigidity are forbidden); ΔR is the rigidity step size in the simulation (Shea et al., 1965; Desorgher, 2005; Desorgher et al., 2006; Herbst

Table 2.2: Properties of some of the HiSPARC stations: geographic longitude (ϕ), geographic latitude (λ), altitude (h), and the geomagnetic vertical cut-off rigidity (R_C) calculated from the PLANETOCOSMICS simulations.

Station Name/ID	R_C [GV]	ϕ [deg]	λ [deg]	h [m]	No. Detectors
Nikhef/501	3.19	4.95 E	52.36 N	56.18	4
College Hageveld/203	3.18	4.63 E	52.35 N	53.71	2
Leiden/3001	3.23	4.45 E	52.17 N	54.08	2
Eindhoven/8001	3.44	5.49 E	51.45 N	70.12	2
Birmingham University/14001	3.06	1.93 W	52.45 N	204.14	4

et al., 2013).

The rigidity spectrum for each of the five HiSPARC stations was investigated to determine R_C for each station. The cut-off rigidity calculated for the five HiSPARC stations for a vertical incidence upon the atmosphere (i.e. 0° zenith angle) are presented in Table 2.2 which show that there is little variation in R_C between the HiSPARC stations and that they observe PCRs with rigidities in excess of ~ 3 GV. This analysis was initially carried out for the vertical direction (i.e. azimuth = 0° , zenith = 0°); however further trajectories were simulated for different azimuth and zenith angles to determine the dependence of the rigidity spectrum on the detector acceptance angle. The analysis for the azimuthal dependence was carried out at a zenith angle of 20° as this is around the most probable angle for HiSPARC events (Fokkema, 2012), and the analysis of the zenith dependence was carried out at an azimuth angle of 0° . This analysis is shown in Figure 2.5, and demonstrates that there is no strong dependence of the azimuth direction or zenith (up to 45°).

The small variation between HiSPARC stations is due to their close proximity in geographic latitude and longitude. The values of R_C calculated for the HiSPARC stations suggest that they should be able to observe higher energy Solar Cosmic Rays (SCRs), but may not be as susceptible as the higher latitude NM where the effects of GLEs are highly observable.

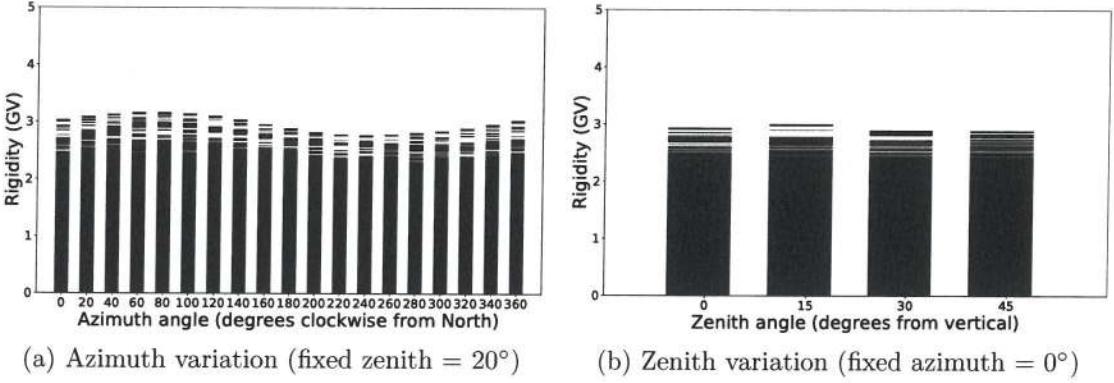


Figure 2.5: Azimuthal and zenith angle variations in the allowed and forbidden rigidity trajectories for HiSPARC station 501 from simulations in steps of rigidity, $\Delta R = 0.01$ GV. The forbidden trajectories are in black; allowed trajectories are in white.

2.4.2 Asymptotic Viewing Directions

Another output from the PLANETOCOSMICS simulations allowed us to understand the directions of moving particles entering the Earth’s magnetosphere prior to their trajectory through the magnetosphere and arrival at the atmosphere. By tracking particle trajectories we can define the Asymptotic Viewing Direction (AVD) of CR stations, which represents the direction of CR motion before entering the magnetosphere and being observed by a detector (McCracken et al., 1962; Danilova et al., 2019). This allowed us to understand the average directions in space that ground-based CR detectors observe. Higher energy CRs are deflected less by the magnetosphere and therefore the AVDs of high rigidity cut-off stations are simply their zenith; however, lower energy CRs are deflected more, hence stations with a lower rigidity cut-off may observe CRs from a range of directions.

It can be seen from Figure 2.6 that the AVDs for each of the HiSPARC stations investigated are rather similar, due to their close geographic proximity, and that they mostly straddle the equator for low rigidity PCRs.

The simulations were performed up to a rigidity of 20 GV, in steps of $\Delta R = 0.01$ GV. The resultant AVDs are limited between $\pm 20^\circ$ latitude. However, at higher rigidities, we would see the AVDs spiral in towards the geographic location of the

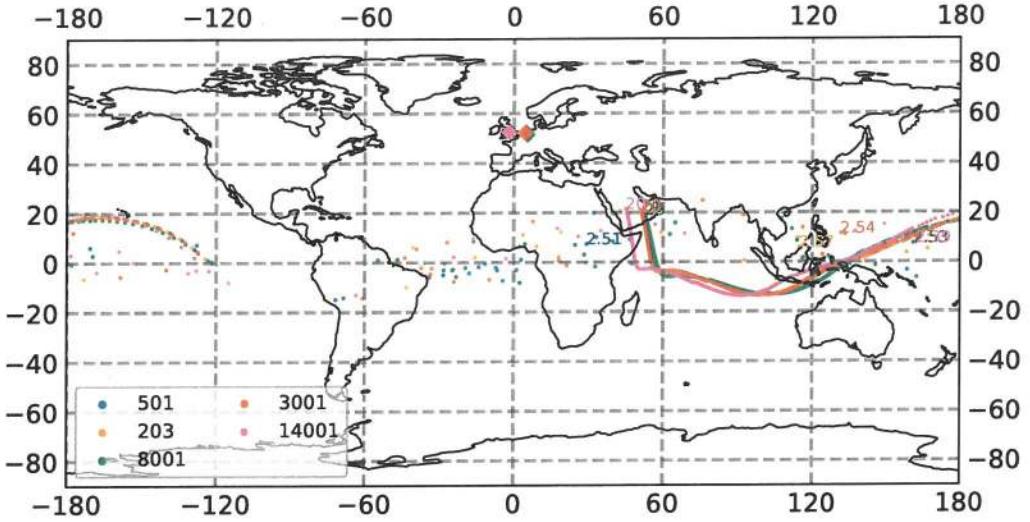


Figure 2.6: The vertical asymptotic viewing directions of 5 HiSPARC stations. The rigidity range of the simulations were from $1.0 \text{ GV} < R < 20.0 \text{ GV}$, and the results are plotted in geographic coordinates on January 20th 2005. The diamonds correspond to the HS ground location and the circles correspond to the AVD for a specific rigidity value.

station and the PCR would enter the magnetosphere and atmosphere almost vertically above the detector. The 20 GV directions are all grouped closely together due to the close geographic locations of the stations. This map of the AVDs also informs us that we should expect to be able to observe some lower energy PCRs when the zenith of the detector is not facing the asymptotic direction of the PCR. The viewing directions of allowed, lower rigidity, trajectories with $R_C \sim 2.5 \text{ GV}$ are shifted eastwardly by $\sim 120^\circ$ longitude. This demonstrates that the allowed, lower-energy PCRs are deflected significantly by the Earth's magnetosphere; we can therefore detect solar eruptive events when the station is not pointing in a direction in-line with the Sun and hence observe SEPs ~ 8 hours before the stations align with the direction of the source.

This has a significant impact on the ability of the HiSPARC detectors to observe transient solar eruptive events which have a SEP spectrum with energies $< 10^9 \text{ eV}$. The ability of HiSPARC to observe highly anisotropic events may be limited because of the similar AVDs for all the HiSPARC stations. For this latter reason, many ground-based CR stations are spread across Earth's surface, to maximise coverage.

2.5 HiSPARC Observations

2.5.1 Observations of Ground Level Enhancements

The search for evidence of GLEs within the HiSPARC data was conducted for the events listed in Table 2.1, as they are the only GLEs that span the operational epoch of the HiSPARC network. Figures 2.7, 2.8, and 2.9 show the HiSPARC observations around GLE 70, 71, and 72, respectively. As highlighted by Fig. 2.4, most of the observations are only the events data (i.e. coincidences between the detectors of a station); however, where possible, we also show the singles rates from each of the individual detectors in a station, when available.

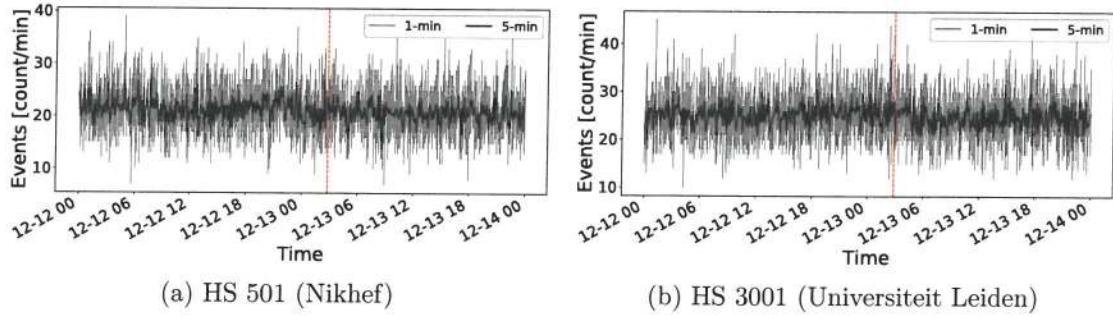


Figure 2.7: HiSPARC data for stations 501 and 3001 around the epoch of GLE 70 on 13/12/2006. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

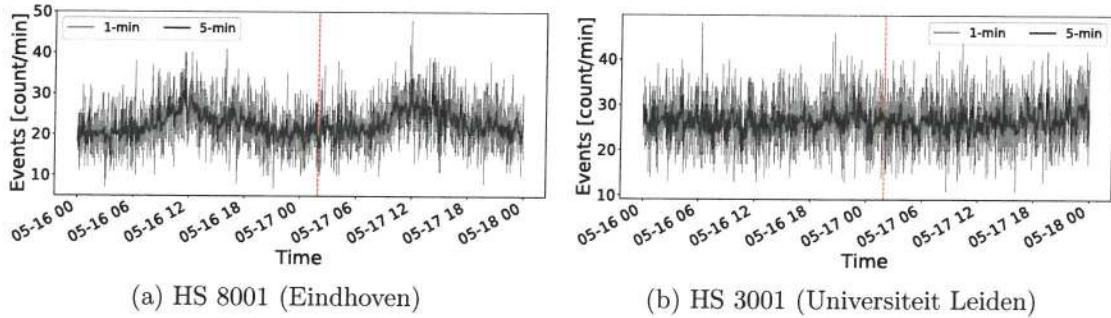


Figure 2.8: HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

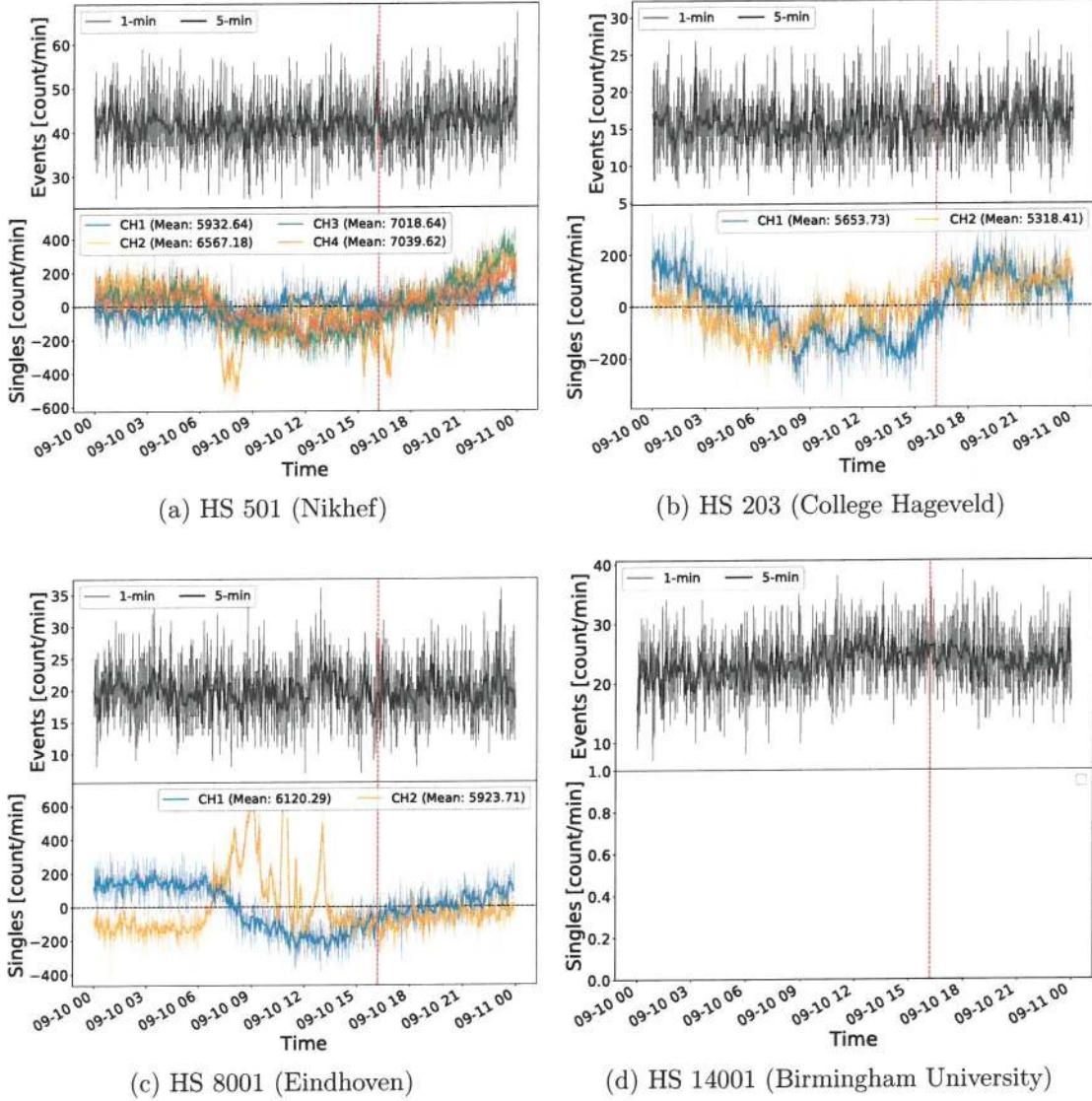


Figure 2.9: HiSPARC data for 4 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the 1-minute (grey) and 5-minute (black) averaged trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH_n) in the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

We can see from Figures 2.7, 2.8, and 2.9 there are no clear and obvious signs of the GLE signals in the HiSPARC observations, as was clear for those given in Fig. 2.1 for the Oulu NM station, in Finland. This is the case for both the events data and the singles data. Furthermore, an anticipated MD increase of $\sim 0.61\%$ – as observed by Timashkov et al. (2008) – was not observed in the HiSPARC data and a GLE with this magnitude is indistinguishable from noise.

There are some excursions from the mean count rate which make it difficult to determine variations from space weather events and other sources; this is significantly more prominent in the singles rates which are shown in the GLE 72 plots (Fig. 2.9) for stations 501, 203, and 8001. These excursions are the effect of atmospheric pressure and temperature on the muon count rates; in Section 2.6 this is discussed further and the effect is accounted for. After its removal we then re-investigated the corrected data, which is discussed in Section 2.7. We also see the existence of a diurnal signal in Figure 2.8a, with CR count rates peaking at around midday. We expect a daily variation from a combination of the CR anisotropy in the interplanetary space and the rotation of the Earth meaning detectors look in different directions over the course of a day. The diurnal variation is expected to have a magnitude of $< 1\%$ (Mishra & Mishra, 2007, 2008; Dubey et al., 2016; Thomas et al., 2017), but here we see an increase of ~ 50 – 100% , which suggests there may be an additional factor causing the signal.

No clear GLEs have been observed in the HiSPARC data. We believe this is due, in-part, to the rigidity cut-off of the HiSPARC stations, as well as the different particle species observed by HiSPARC compared to the NMs. However, Humble et al. (2012) state that NM stations with cut-off rigidities up to ~ 15 GV observed the GLE in September 1989, indicating that SEPs up to at least ~ 15 GeV must have been present in the spectrum during that particularly large event, and similarly up to ~ 30 GV was reported by Lovell et al. (1998). Therefore, the rigidity cut-off of the HiSPARC stations is not necessarily the limiting factor. It is also possible that the SEPs which induced these GLEs were lower-energy PCRs and therefore were insufficient to produce EASs of muons; we know the HiSPARC network is nominally used for observations of UHECRs with energies several orders of magnitude larger.

GLEs are normally associated with SEPs with energies in the MeV to low-GeV regime; hence why GLEs are typically observed by NMs. Only the most energetic events have been observed by MDs as they are more sensitive to the hard component

of the PCR spectrum (Augusto et al., 2016). Observations of GLEs with HiSPARC will therefore only be linked with the maximum energy release during solar flares, and often this is very short and still the SEPs are of insufficient energy to be detected by MDs (McCracken et al., 2012; Augusto et al., 2016).

To further understand the impact of SEP energies on the HiSPARC observations, in Section 2.8 we used air shower simulations to investigate the CR spectrum at quiet times and during GLEs.

2.5.2 Observations of Forbush Decreases

The search for evidence of FDs within the HiSPARC data was conducted for the FDs highlighted in Table 2.1. Figures 2.10, 2.11, and 2.12 show the HiSPARC observations around the epochs of the first four FDs listed in Table 2.1. Each of the plots shows only observations using the HiSPARC events data (i.e. coincidences between the detectors of a station), as singles data were not available at those epochs (see Fig. 2.4).

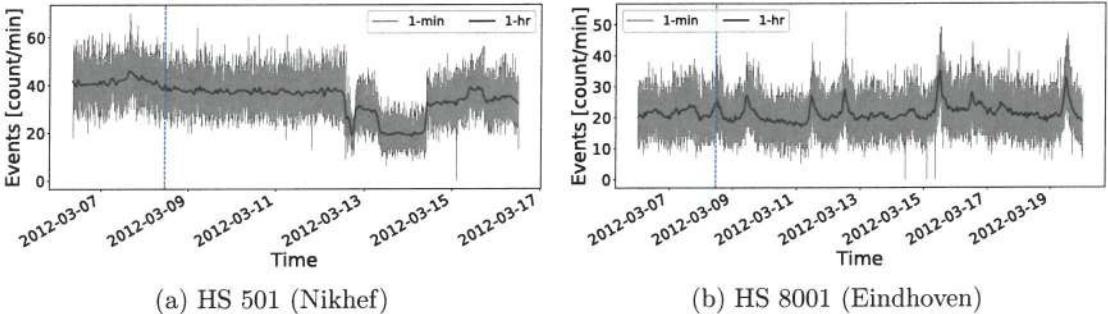


Figure 2.10: HiSPARC data for stations 501 and 8001 around the epoch of the FDs in March 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed lines show the approximate onset-time of the FDs. The units of time on the x-axis are, YYYY-MM-DD.

We can see from the plots that there were no clear signs of the FD signals in the HiSPARC data shown here. We observed a set of significant decreases in the muon count rate in station 501 after the second FD in March 2012 (see Figure 2.10a); however, it is unclear whether this was a consequence of the FD or other hardware

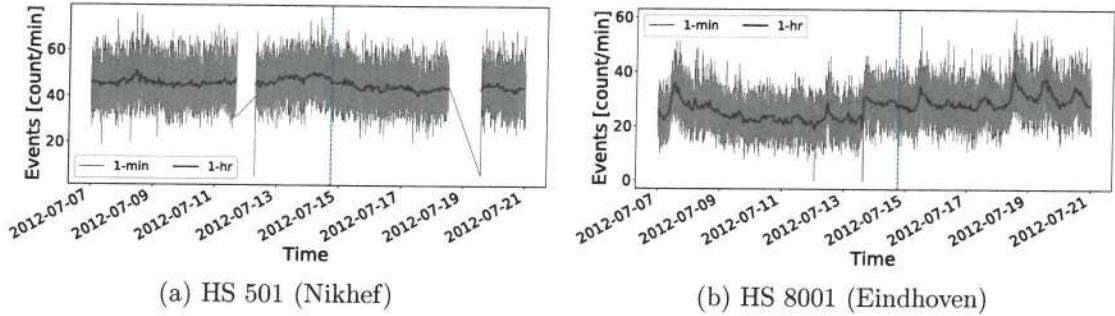


Figure 2.11: HiSPARC data for stations 501 and 8001 around the epoch of the FD in July 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

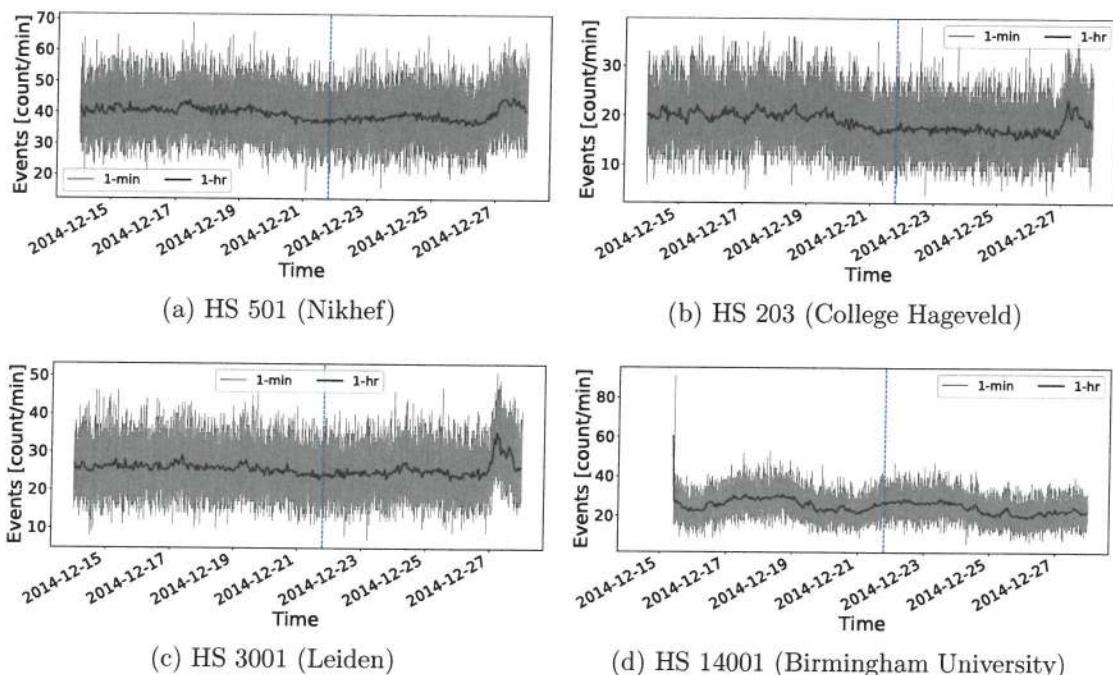


Figure 2.12: HiSPARC data for four stations around the epoch of the FD in December 2014. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

reasons, as the FD was not observed in the other HiSPARC stations. The shape of the FD in the NM data shows a sudden decrease and a smooth recovery within two days, but the shape of the HiSPARC observations traces a more complex evolution, which suggests that the cause is not the FD, but rather a result of hardware.

As we saw with the earlier investigation into GLEs, we again see excursions from

the mean count rate which vary over longer time scales. This is due to metrological variations in the atmosphere. In Section 2.6 this is discussed further and the effect is accounted for.

In addition, it is quite clear from Figure 2.10b and Figure 2.11b that station 8001 (Eindhoven) displays a semi-persistent diurnal variation in the count rate. We typically expect a diurnal variation of $< 1\%$ (Mishra & Mishra, 2007, 2008; Dubey et al., 2016; Thomas et al., 2017), but here we see an increase of $\sim 100\%$, which again suggests there may be an additional component. *suggestion of what?*

For the final two FDs listed in Table 2.1, the HiSPARC observations are shown in Figure 2.13. Plotted are the HiSPARC events data; however, where possible, we also show the singles rates from each of the individual detectors in a station, when available. Furthermore, as these FDs were precursory to GLE 72, we also marked on the epoch of the GLE for completeness.

As with the other FD epochs, we again do not observe any clear signs of the FD signals in either the events nor singles data. In each of the three stations for which there were singles data, we observed a semi-persistent diurnal signal. This was also seen in the events data for station 14001. In the singles data we see a strong diurnal variation of up to $\sim 100\%$ (see channel 2 in Fig. 2.13c). As previously discussed, we expect a diurnal variation of $< 1\%$, increasing around local midday (Mishra & Mishra, 2007, 2008; Dubey et al., 2016; Thomas et al., 2017); however, in Figure 2.13c not only do we see a stronger variation, we also see that the two detectors are anti-correlated, suggested that this signal is a manifestation of thermally induced noise in each detector.

We conclude that no clear signature of FDs has been observed in the HiSPARC data. We again believe this could be linked to the rigidity cut-off of the HiSPARC stations, but also it is possible that the reason originates in the lower sensitivity to lower rigidity PCRs of MDs compared to NMs. We also note that the atmospheric effects in the raw data limit our ability to observe the space weather effects, therefore

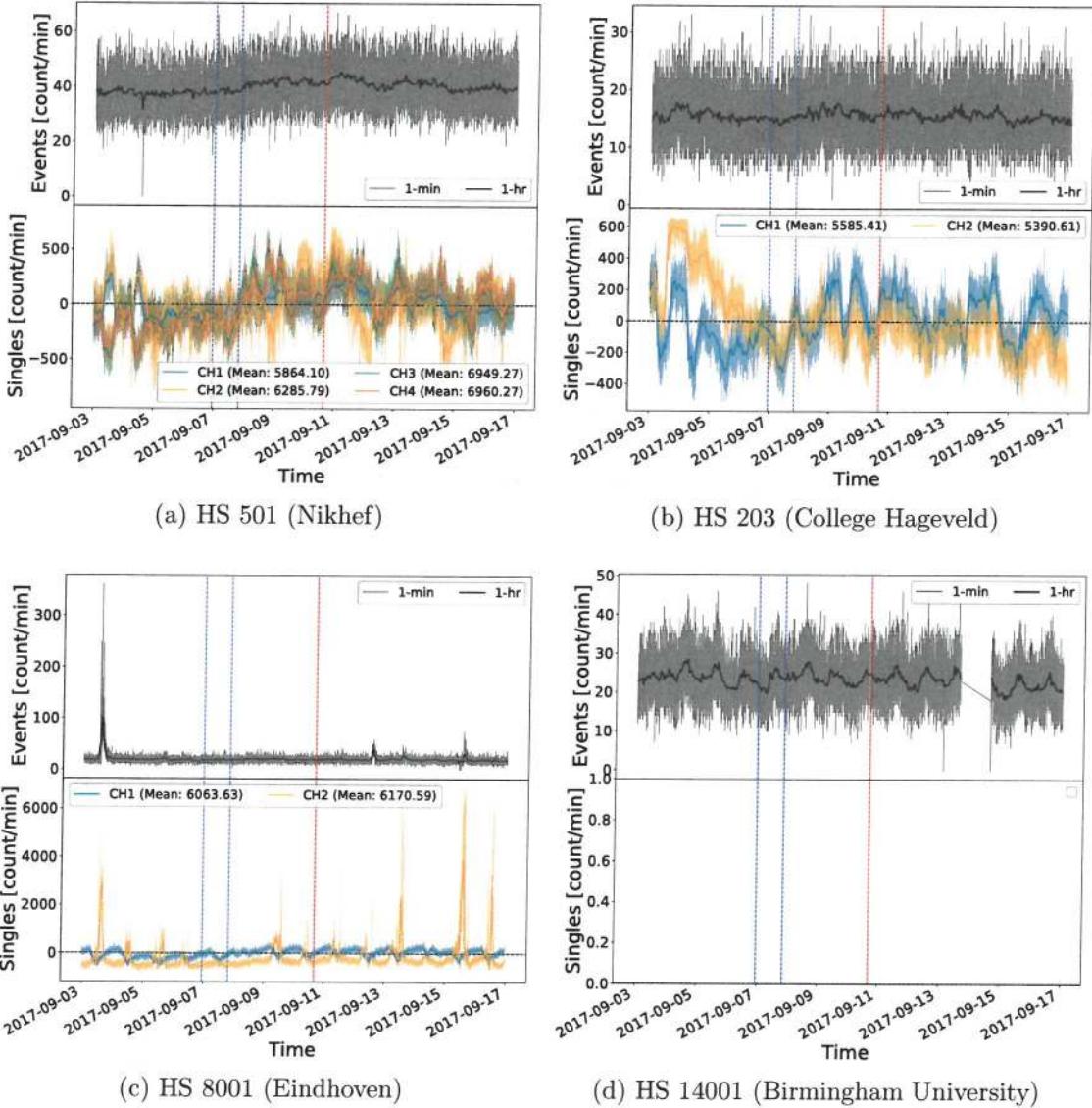


Figure 2.13: HiSPARC data for four stations around the epoch in which there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station, while the bottom panel shows the minute- and hourly-averaged singles counts, mean-subtracted, for each individual detector (or channel, CH_n) in the station. The vertical blue-dashed lines show the approximate onset-times of the two FDs observed around this epoch and the red-dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, YYYY-MM-DD.

in the next section we remove these effects to standardise the HiSPARC data.

I don't feel like this was demonstrated in 2.5, specifically that the excursions were due to atmospheric variations.

Add some further description in 2.5

2.6 Atmospheric Corrections of HiSPARC Data

2.6.1 Motivation

It is well known that observations made by ground-based CR detectors are susceptible to atmospheric conditions (Dorman, 2004a, 2010; Berkova et al., 2011; De Mendonça et al., 2013; Paschalis et al., 2013). As we have seen in the plots shown in Section 2.5, there exist excursions in the data whose origins are from atmospheric variations. This sensitivity makes it difficult to differentiate between variations due to space weather events and those due to Earth's atmospheric conditions, therefore it was necessary to correct for these effects in the data.

2.6.2 Barometric Correction

Atmospheric pressure affects the CR path length due to the expansion and contraction of the atmosphere with varying pressure (Dorman, 1972; Paschalis et al., 2013); hence the CR counts are observed to be negatively correlated to atmospheric pressure as shown for both NMs and MDs in Figure 2.14.

A correction for this barometric effect is routinely applied as part of the data calibration for all NM stations within the NMDB and an online barometric coefficient tool⁴ is available for NMs, which allows users to perform the barometric correction for a given station over a user-defined epoch (Paschalis et al., 2013). There is no such process routinely applied in the HiSPARC data pipeline.

The method of correcting for the barometric effect is discussed widely in the literature regarding NMs and is shown to depend on the barometric coefficient (Paschalis et al., 2013). Assuming the cosmic ray flux variation, absent of the atmospheric effects, is reasonably stable, then a simple correction to the counts can be made. The

⁴<http://cosray.phys.uoa.gr/index.php/data/nm-barometric-coefficient>

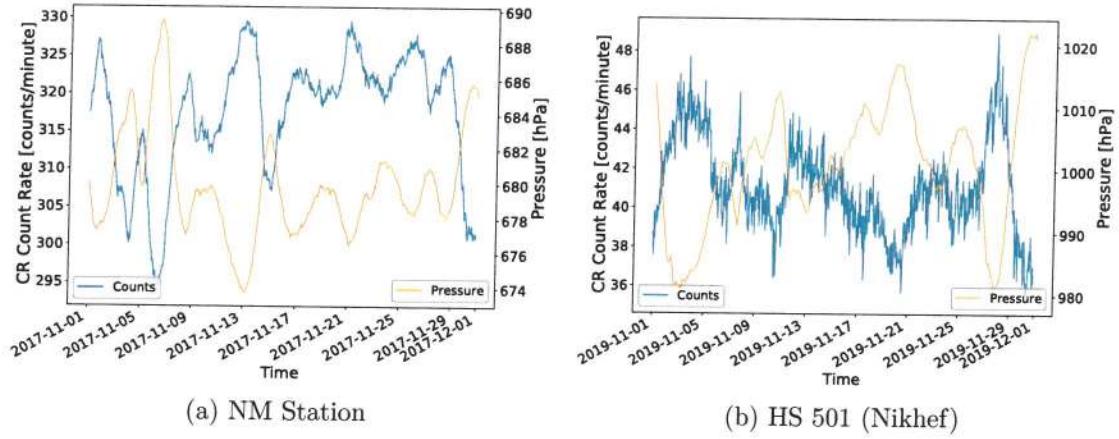


Figure 2.14: The anti-correlation between CR count rates and the atmospheric pressure. (a) shows the CR and the local atmospheric pressure measured at a NM in the South Pole; (b) shows the CR and pressure measured by HiSPARC station 501. In both plots, the data shown are hourly-averaged, to highlight the effects. The units of time on the x-axis are, YYYY-MM-DD.

CR variations that depend on the local atmospheric pressure are described by:

$$\Delta N = -\beta N \Delta P, \quad (2.2)$$

where ΔN is the change in count rate, N is the count rate, β is the barometric coefficient, and $\Delta P = P - P_0$ is the deviation in pressure from the average (P_0) in the given time-period (Paschalis et al., 2013)

Through the integration of equation (2.2), the solution shows the dependence of cosmic ray intensity on pressure,

$$N = N_0 e^{-\beta \Delta P}. \quad (2.3)$$

Therefore by taking the logarithm of equation (2.3), one can obtain the barometric coefficient by fitting the linear model to the observed data, of the form:

$$\ln \left(\frac{N}{N_0} \right) = -\beta \Delta P, \quad (2.4)$$

where N_0 may be considered as the mean count rate over the given time-period of

observations.

A demonstration of this method is shown for both a NM and a HiSPARC station in Figure 2.15. In both cases the linear fit does a good job of finding the barometric coefficient and was used to remove the pressure effect from the data.

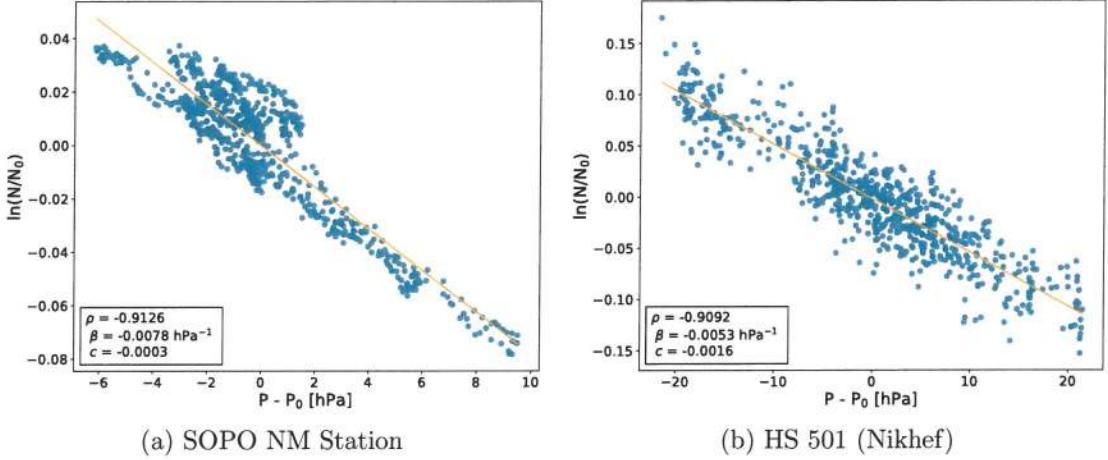


Figure 2.15: The barometric coefficient calculation: (a) during November 2017 for the South Pole (SOPO) NM station, (b) during November 2019 for HiSPARC station 501 at Nikhef.

For comparison, and to show the success of this method at removing the pressure variation, the raw and corrected HiSPARC data are shown in Figure 2.16. It is clear from Figure 2.16 that the large excursions are adequately removed from the data after the correction.

Using the online barometric coefficient tool (Paschalis et al., 2013), it was possible to also provide a comparison between the method used in this work and the correction of the NMDB stations, as a further validation. This is shown in Figure 2.17 for monthly corrections throughout 2017 for the NM station at the South Pole (SOPO).

Figure 2.17 shows a close agreement between the barometric coefficient calculated in this work and those acquired from the online tool for the SOPO NM. This was also true for other stations tested (Apacity and Rome), thus validating that the method used in this work was suitable for application on the HiSPARC data. The barometric

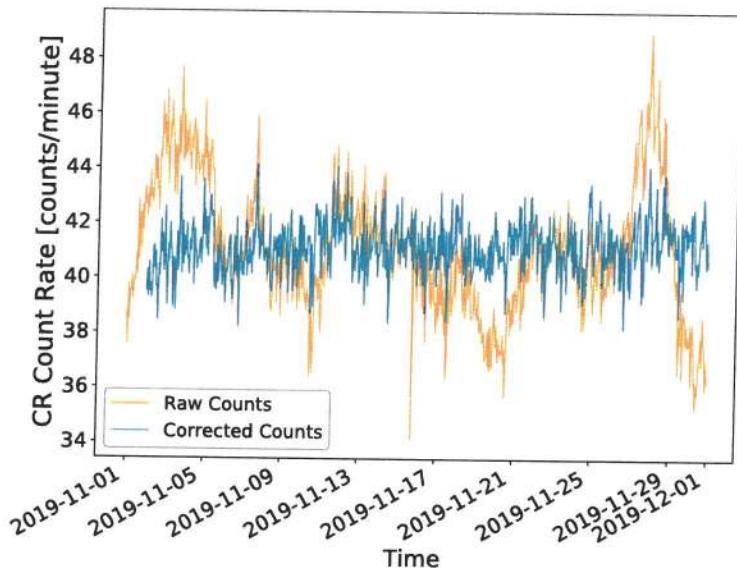


Figure 2.16: A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the pressure correction (blue line). The units of time on the x-axis are, YYYY-MM-DD.

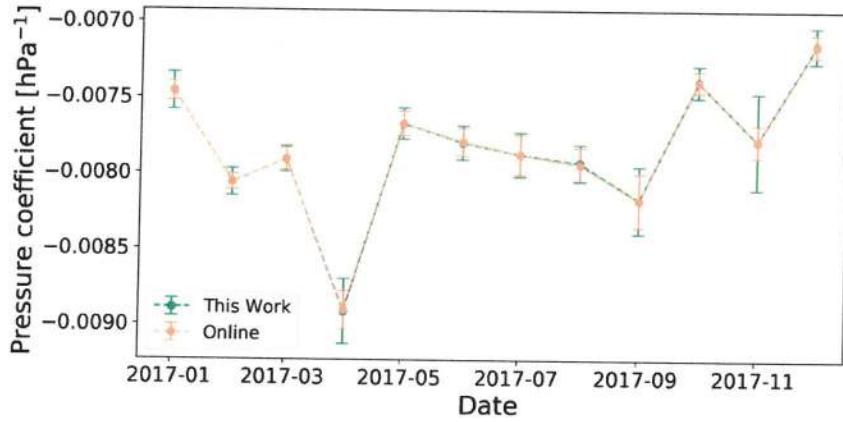


Figure 2.17: A comparison between the monthly barometric coefficient computed in this work and using the online barometric coefficient tool throughout the year 2017 for the SOPO NM station. The units of time on the x-axis are, YYYY-MM.

correction was performed on the HiSPARC data for stations where sufficient pressure data and count rates existed.

2.6.3 Temperature Correction

The effects of atmospheric temperature have been shown to influence both the creation and disintegration processes for muons in the atmosphere (Berkova et al., 2011). There is hence a positive effect and a negative effect on muon intensity as a

consequence of temperature variations (Mendonça et al., 2016).

The positive effect is related to pion decay and its dependence on temperature variation. The higher the temperature, the lower the atmospheric pion absorption, which implies a higher generation rate of muons (Mendonça et al., 2016).

The negative effect corresponds to the decrease of muon intensity at ground level as the muon average path length varies with temperature. Due to the heating and expansion of the atmosphere during summer periods muons are produced higher in the atmosphere; hence the muon propagation path length increases meaning there is more atmosphere for muons to traverse before reaching the ground, and an increased decay probability and ionisation losses (Savić et al., 2015; Mendonça et al., 2016).

Due to the difference in decay probability, the negative effect dominates for low energy muons and the positive effect dominates for high energy muons (Berkova et al., 2011). It is therefore expected that the negative effect should dominate for the HiSPARC network. This is in contradiction with the observations of diurnal variation with the HiSPARC detectors, as one can quite clearly see that the HiSPARC stations register higher count rates during local noon (see Fig. 2.9c). However, when observing the singles rates, we do see some detectors displaying a positive effect and some displaying a negative effect (see Fig. 2.9). This is not consistent between stations, and provides more evidence to suggest this is an effect of thermally induced noise in the Photo Multiplier Tubes (PMTs).

Several methods of correcting for the temperature effect are discussed in the literature, e.g. see Berkova et al. (2011) and Mendonça et al. (2016) for a summary. However, the methods discussed are typically applied over ~~100~~ timescales of years, to account for seasonal variations with low temporal resolution, rather than to account for short timescale variations with periods of less than a day ~~100~~ T . These methods are not necessarily suitable for this work on ephemeral space weather events.

The HiSPARC stations provide the local outdoor temperature which is measured nearby the detectors and also the temperature in the room where the electronics are

what is the correlation value? ~~Please~~ Include this.

located. The latter is not of use, but the former can be used for temperature corrections. Figure 2.18a shows the pressure corrected events data and the temperature data for station 8001 in July 2012. There is a strong correlation between the events and the temperature, which is demonstrated in Figure 2.18b, yet weaker than with pressure.

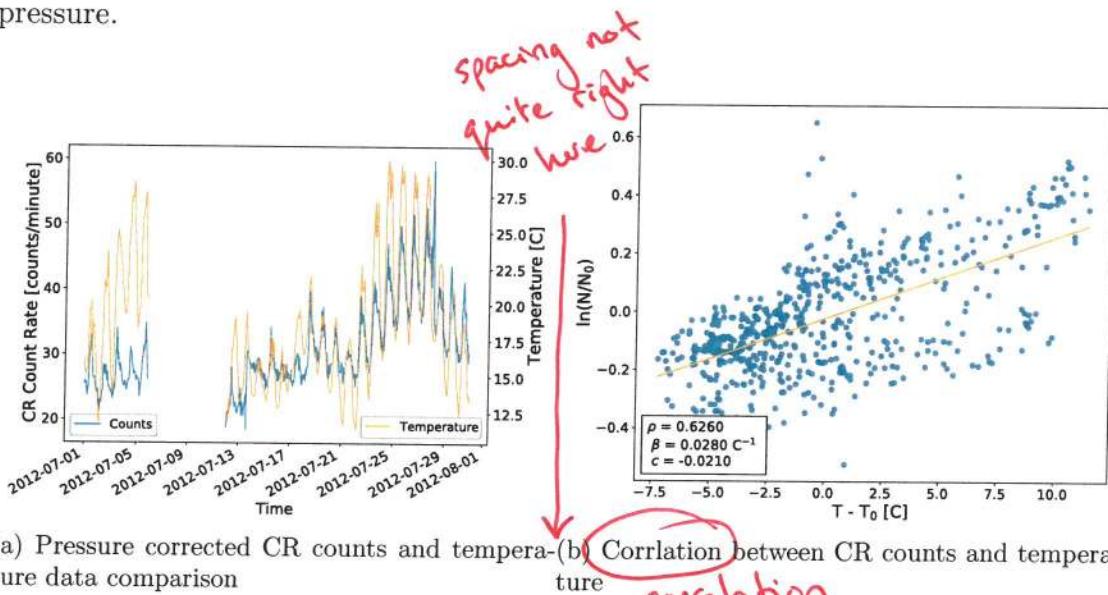


Figure 2.18: The relationship between the pressure corrected events data and the outdoor temperature as measured at HiSPARC station 8001 (Eindhoven). (a) shows the time-series of hourly-averaged pressure corrected events and temperature data; (b) shows the correlation between the counts and temperature, and the fitted line to calculate the correction coefficient.

We simplified the correction method and used a linear fitting technique (i.e. the same as the barometric correction in Section 2.6.2; ~~however~~, replacing the pressure for locally measured outdoor temperature). We can see from Figure 2.19 that this correction method does remove a significant amount of the variation correlated with the outdoor variation, including dampening the strong diurnal variation. There still persists some strong variation in the count rate however, which shows that this method is not as effective as the pressure correction. Nevertheless, we continue with this temperature correction as it is sufficient at reducing the diurnal variation in the data around the events considered.

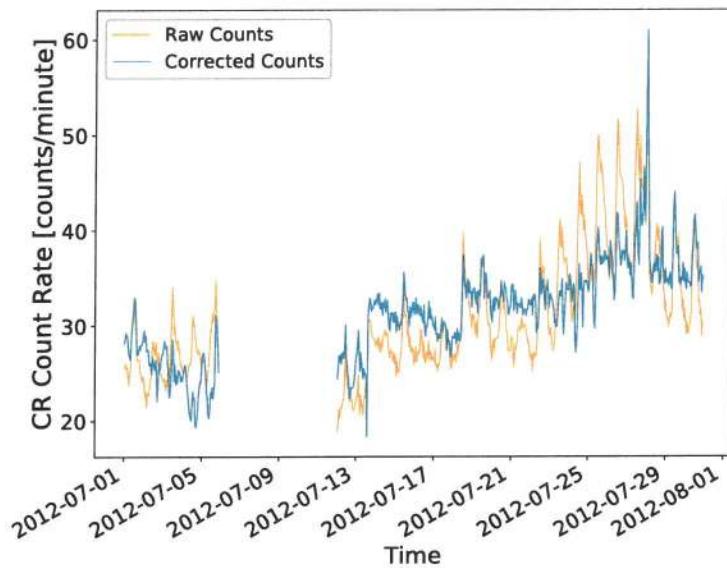


Figure 2.19: A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the temperature correction (blue line). After the correction, the diurnal variation from the temperature effect has been reduced.

2.7 HiSPARC Observations After Atmospheric Corrections

2.7.1 Observations of Ground Level Enhancements

Again, I would find the dates / years at least of these events useful.

Following the atmospheric correction, the search for evidence of GLEs was repeated, this time within the corrected HiSPARC data. This could only be done for GLE 71 and 72, as the HiSPARC network was not collecting meteorological data during the epoch of GLE 70. Figure 2.20 and Figure 2.21 show the atmospheric-effect corrected HiSPARC observations around the epochs of GLE 71 and 72, respectively. The observations of GLE 71 show only the HiSPARC events data; however, we also show the singles rates from each of the individual detectors in a station for GLE 72.

Despite the atmospheric correction, in general, doing a good job at removing atmospheric variations in the CR counts, there still remained no clear GLEs observations in the pressure corrected HiSPARC data. We believe this is due to a mixture of the reasons discussed above. A high rigidity cut-off of the HiSPARC

Could you, at least in brackets, mention what specifically you are referring to

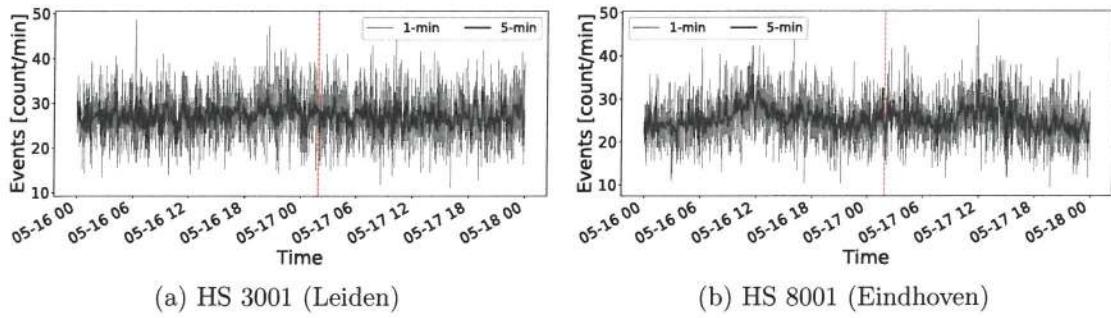


Figure 2.20: Atmospheric-corrected HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

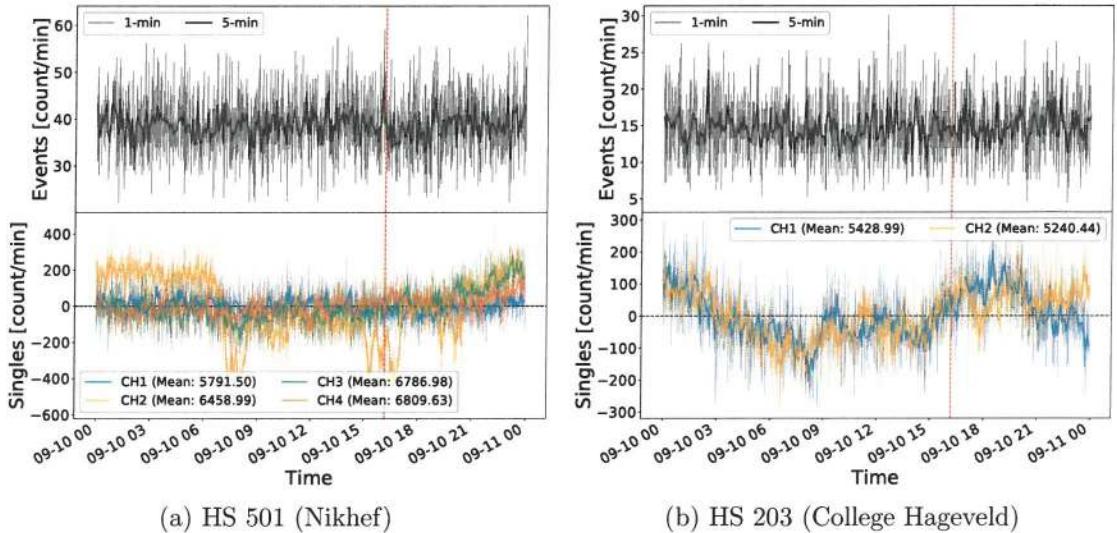


Figure 2.21: Atmospheric-corrected HiSPARC data for 2 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the minute-averaged (grey) and 5-minute-averaged (black) trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH_n) in the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

stations leads to a low increase in the CR count, as GLEs are typically caused by SEPs with a lower energy; hence too few additional muons were produced during the GLEs. Again we should however note that Humble et al. (2012) and Lovell et al. (1998) stated that NM stations with cut-off rigidities up to \sim 15–30 GV observed the GLE in September 1989, indicating that high-energy SEPs may be present and

could you provide some comparison to how much “bigger” this storm was compared to the ⁷⁶ones you are considering.

also cause GLEs, suggesting the rigidity cut-off may not be the limiting factor. In addition, as discussed above, the events data require a trigger between at least 2 detectors that are separated by up to 10 m, therefore this further biases the stations to be sensitive to more energetic PCRs. Furthermore, the issue could be due to the particle species, re-emphasising the wide use of NMs rather than MDs.

This provides motivation to investigate the flux of muons at ground level during quiet periods, i.e. from GCRs, compared to the flux of muons at ground level during energetic solar events such as GLEs. This work was performed and is discussed in Section 2.8.

2.7.2 Observations of Forbush Decreases

The search for evidence of FDs was also re-conducted using the atmospheric-effect corrected HiSPARC data. Figure 2.22 and Figure 2.23 shows the corrected HiSPARC observations around the epochs of a FD in July 2012 and December 2014, respectively.

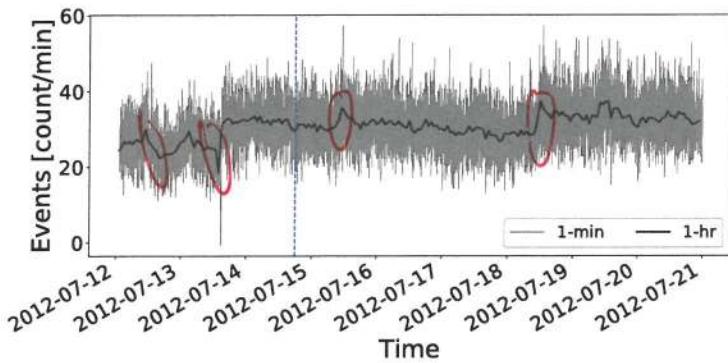


Figure 2.22: Atmospheric-corrected HiSPARC data for station 8001 (Eindhoven) around the epoch of the FD in July 2012. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

In Figure 2.22 there are no clear FD observations in the corrected HiSPARC data; however, in Figure 2.23 there is a slight indication of a $\sim 2\%$ decrease in the count rate at the epoch of the expected FD. This suggests that after correcting the data for atmospheric effects, that FDs on the order of $\sim 2\%$ might be observed by

I can't see any obvious change. What makes you think this 2% isn't just random?
Especially compared to bigger jumps in Fig 2.22.

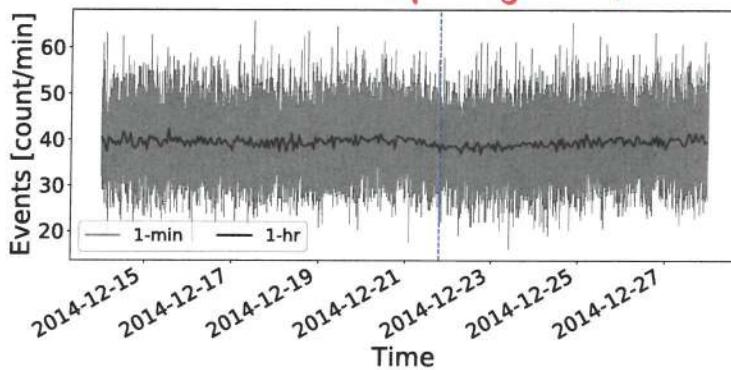


Figure 2.23: Atmospheric-corrected HiSPARC data for station 501 (Nikhef) around the epoch of the FD in December 2014. The plot shows the minute-averaged (grey) and hourly-averaged (black) trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

the HiSPARC network.

Finally, the atmospheric-effects corrected observations of the two FDs which occurred around GLE 72 is shown in Figure 2.24, showing the corrected events data and corrected singles data for each of the individual detectors in the stations.

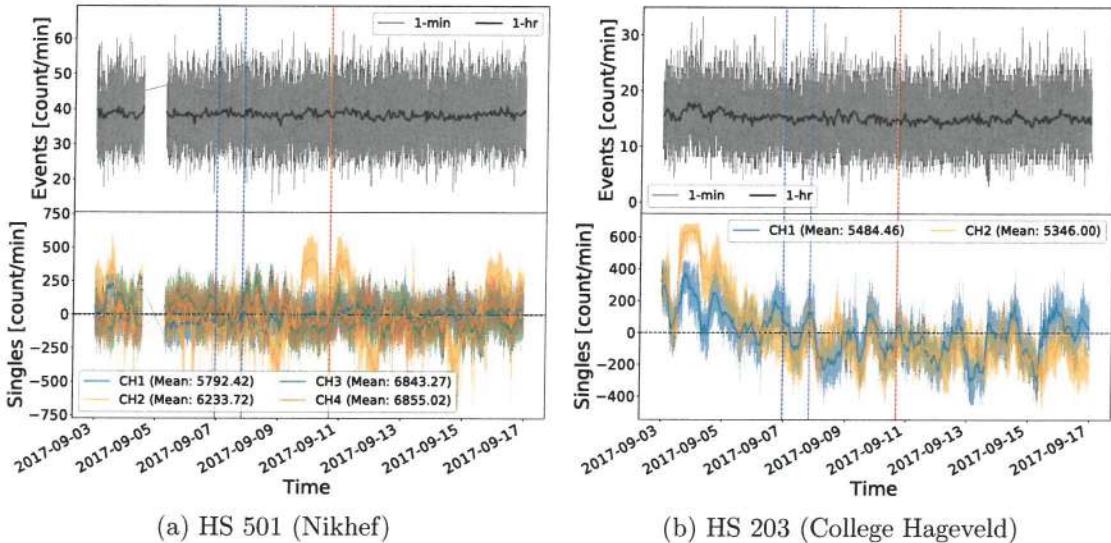


Figure 2.24: Atmospheric-corrected HiSPARC data for 2 stations in an epoch where there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute-averaged (grey) and hourly-averaged (black) trigger events. The bottom panel shows the minute- and hourly-averaged singles counts, mean-subtracted, for each individual detector (or signal channel, CH_n) in the station. The vertical blue-dashed lines show the approximate onset-times of the FDs and the red-dashed line depicts the approximate onset-time of the GLE. The units of time on the x-axis are, YYYY-MM-DD.

There are no clear FDs in Figure 2.24, and the singles plots demonstrate that the atmospheric corrections do not sufficiently remove all the signal excursions from the mean, and there remain large (up to 10%) deviations in the signals. One reason for this, particularly for the singles, is that the measured temperature is recorded in the atmosphere and not inside the roof boxes containing the PMTs. The roof boxes are made of plastic and one would suspect the air temperature inside the boxes is different to the ambient atmospheric temperature measured nearby. In order to fully remove the effects of thermally induced count variations, we must record the more accurate temperatures of the PMTs instead of just the atmospheric temperature.

2.8 Air Shower Simulations

2.8.1 Motivation

There was no evidence to suggest the GLEs were observed in the HiSPARC data, even after correcting for atmospheric effects of pressure and temperature. This leads us to question whether it is possible to observe GLEs with the HiSPARC detectors. In order to answer this, we needed to understand the muon flux at ground level and the scale of air shower muon footprints produced by PCR. To investigate this, simulations of air shower development were performed for a range of PCR energies for both primary protons and α -particles.

To simulate the CR air shower development, the Cosmic Ray Simulations for Kascade (CORSIKA) software was employed: a Monte Carlo programme providing detailed simulations of the evolution of air showers initiated by PCR through the atmosphere (Heck & Pierog, 2017). The particles in the CORSIKA simulations are tracked through the atmosphere until they undergo interactions with atmospheric nuclei, decay due to their instability, or reach the ground level defined as the simulation terminator.

Proton and α -particle initiated air showers were generated with energies ranging from 10^9 to 10^{20} eV, and 4×10^9 to 10^{20} eV, respectively. In total $\sim 2 \times 10^5$ proton-

(and in fact you use 'to' in Appendix A)

im only suggesting this change because 4×10^9 - 10^{20} looks more like an equation than range.

initiated showers were simulated and $\sim 2 \times 10^5$ α -particle-initiated air showers were simulated. Lists detailing the breakdown of PCR energies and number of simulations is provided in Appendix A, along with a brief discussion of the settings chosen within the simulations.

2.8.2 Air Shower Footprints

The average footprint of the muons at ground level, due to PCRs, was calculated from the output of the CORSIKA simulations. This was achieved by taking the distribution of the number of muons at ground level at the end of the simulations as a function of their radial distance from the shower core, as this distance was provided as an output from the simulations. Multiple realisations of the air showers were simulated (see Table A.1). For a given PCR energy, the mean distribution of radial footprints was calculated by averaging over all of the individual simulations. Figure 2.25 shows the radial distribution of muons at ground level for air showers induced by vertically incident protons and α -particles.

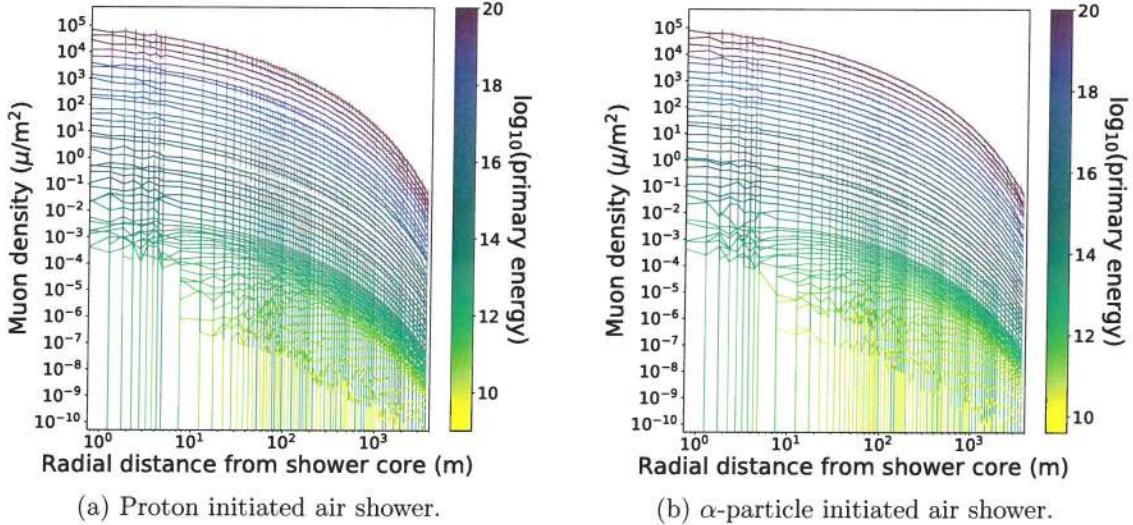


Figure 2.25: Mean muon density footprints for (a) proton-initiated air showers and (b) α -particle-initiated air showers with initial PCR trajectories with zenith angles $\theta = 0^\circ$ and various PCR energies. The error bars given represent 1σ .

In addition to the vertically induced air showers, we also repeated the simulations for air showers randomly selected from a uniform distribution of incident

angles between 0° (vertical) and 70° , to provide a simulation that is more physically representative of CRs arriving from all directions. Radial distributions of muons were produced, similar to those in Figure 2.25, but they are not shown here as the difference is not drastically different by-eye.

The interpretation of Figure 2.25 provides an understanding of the minimum energy PCRs observable by the HiSPARC network. The typical separation between the detectors in a HiSPARC station is ~ 10 m; however, the separation between detectors varies from station-to-station and can be up to as much as 20 m or as low as just a couple of metres. To trigger two HiSPARC detectors, that are separated by approximately 10 m, by at least two muons from the same shower, the density of the shower needs to be at least $\sim 0.005 \mu\text{m}^{-2}$. The simulations therefore suggest that HiSPARC stations typically observe PCRs with a minimum energy on the order of $\sim 10^{13} - 10^{14}$ eV, which agrees with what was found by van Dam et al. (2020b).

This helps to explain why the GLEs and FDs were not observed in the HiSPARC events data. The effects of GLEs and FDs are more prominent at lower PCR rigidities (Belov et al., 2005) and we showed here the air showers induced by these particles are not sufficient to induce an air shower that will trigger multiple detectors in a station. We see that lower energy PCRs do not produce EASs, but rather a very diffuse scattering of muons reach ground level. It is therefore clearer why we did not observe any GLEs in the events data, as the SEP-induced muons are insufficiently spread to trigger multiple detectors in coincidence. It would have been more likely to have observed GLEs in the singles data, as this only records the count rate of an individual detector, which has been shown to have a high muon-detection efficiency close to 100% (van Dam et al., 2020b). However, again the space weather events were not observed in the HiSPARC singles data, which may be further explained by instead investigating the flux of muons at ground level.

2.8.3 Muon Flux

Another output from the CORSIKA simulations was the energy of the muons that reach ground level. From the air shower simulations it was therefore possible to compute an estimate of the energy distribution of muons produced per PCR. Figure 2.26 shows the energy distribution of muons produced per primary PCR, for air showers induced by vertically incident protons and α -particles. The vertically incident air showers provide an upper boundary on the muon flux, but we also repeated the simulations for air showers randomly selected from a uniform distribution of incident angles between 0° (vertical) and 70° , to provide a more physically representative flux. Similar plots were produced to those in Figure 2.26, but they are not shown here, as the difference is again not drastically different by-eye.

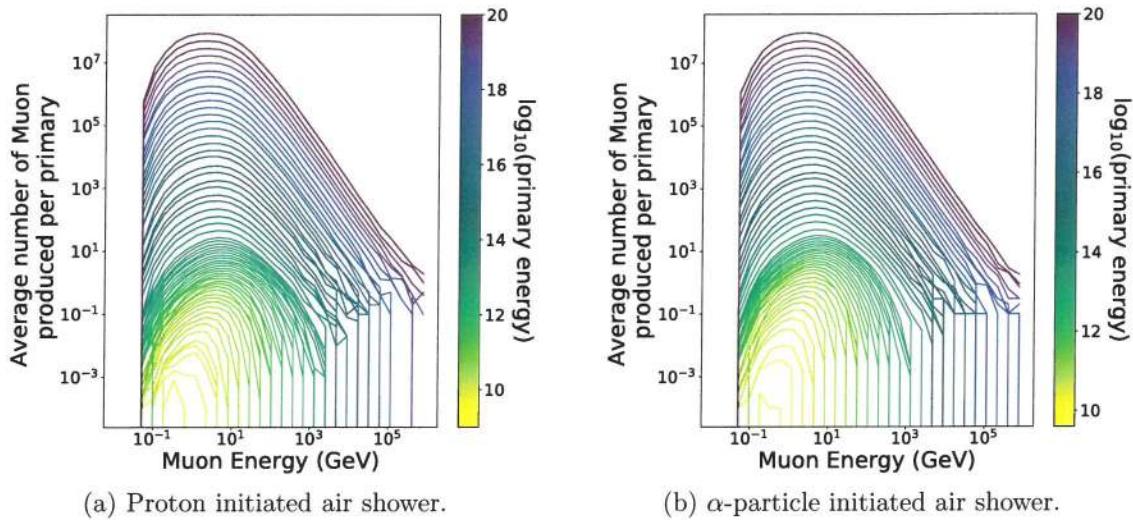


Figure 2.26: Mean number of muons produced at ground level from (a) proton-initiated air showers and (b) α -particle-initiated air showers, with initial PCR trajectories with zenith angles $\theta = 0^\circ$ and various PCR energies.

We see from this analysis that PCRs with an energy less than $\sim 10^{11} - 10^{12}$ eV produce on the order of only one muon that reaches ground level, and below this energy it is rare that any muons arrive at ground level. Knowing the effects of GLEs and FDs are more prominent at lower PCR rigidities (Belov et al., 2005), i.e. PCR energies $< 10^9$ eV, this helps explain why the space weather events were not

observed in the HiSPARC events data. Air showers induced by the lower rigidity PCRs are insufficient to produce significant variations in the flux of the muons at ground level, thus we do not observe a variation from the typical SEPs that induce GLEs.

We also used the data from the simulations to estimate the total muon flux at ground level, based on the PCR flux at the top of the atmosphere. We used a model for the GCR flux at Solar Maximum (Corti et al., 2019), which utilised measurements from the Alpha Magnetic Spectrometer (AMS-02) on-board the International Space Station (ISS) to estimate the flux of PCRs at the top of the atmosphere. Figure 2.27 shows the computed differential flux of muons at ground level, based on the simulations of vertically incident PCRs and those randomly simulated within a 70° acceptance cone. From Figure 2.27, we see that the ground-based flux is similar for both types of simulation performed. In both, the low-energy muon flux dominates and peaks at a muon energy of ~ 1 GeV.

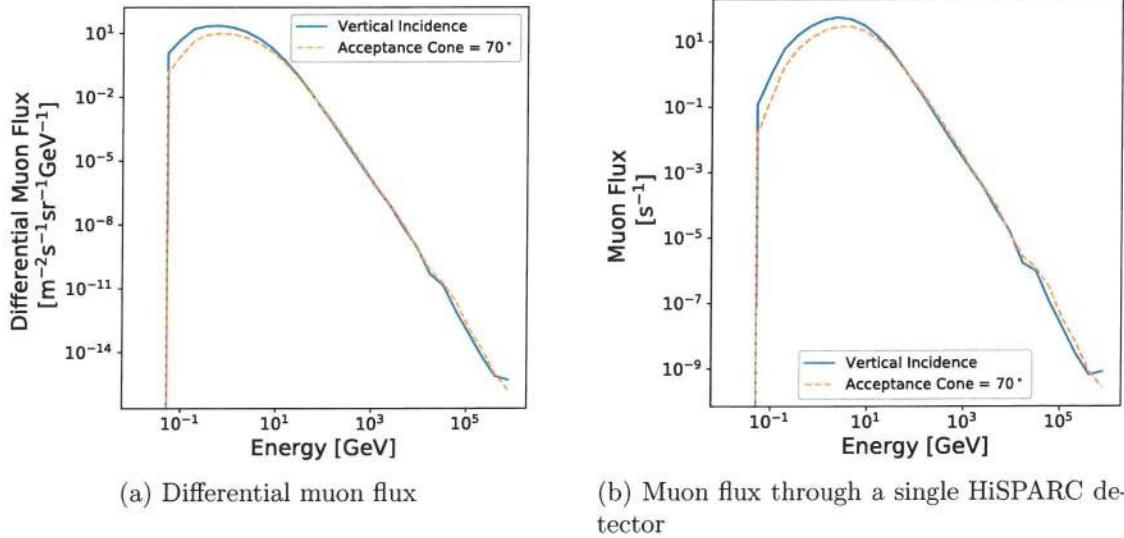


Figure 2.27: Ground level muon spectra as computed in the CORSIKA simulations. (a) shows the differential muon flux at ground level; (b) shows the muon flux through a single HiSPARC detector. In both plots the solid, blue line shows the simulations using vertically incident PCRs, and the dashed, orange line shows the simulations using PCRs incident within a cone of 70° .

The calculated spectra were used to determine the expected rate of muons passing

through a single HiSPARC detector. We computed the rates: $\sim 85 \mu\text{s}^{-1}$ (for non-vertical simulations, i.e. 70° acceptance cone), and $160 \mu\text{s}^{-1}$ (for vertical simulations). These rates are comparable to the generally accepted, average ground level muon flux on the order of $\sim 70 \text{ m}^{-2}\text{s}^{-1}\text{sr}^{-1}$ (Cecchini & Sioli, 2000; Blackmore et al., 2015; Pereira et al., 2021; Particle Data Group et al., 2020), or $1\text{--}2 \mu\text{cm}^{-2} \text{min}^{-1}$ (Particle Data Group et al., 2020), which imply a rate of $\sim 140 \mu\text{s}^{-1}$ through a single HiSPARC detector.

To understand, from the CORSIKA simulations, how the muon count rate measured by the HiSPARC stations changes during a GLE, the PCR spectrum during the space weather events is needed. Unfortunately this information was not so easily acquired; however, there is a tool which provided muon fluxes based on the GCR spectrum and also GLE spectra, which is described in the next section.

2.8.4 Muon Flux From MAIRE

As a further comparison, to investigate how the muon flux varied during GLEs we used the online Model for Atmospheric Ionising Radiation Effects (MAIRE) tool to compute the muon spectrum at ground level (Dyer et al., 2003; Lei et al., 2004). MAIRE allows the computation of the secondary particle spectra in the atmosphere, caused by SEPs, including the ground level neutron and muon fluxes. MAIRE has an advantage of also having the PCR spectra for a number of GLEs built-in; however, they are for the strongest GLEs to-date, which we need to take into consideration, so we do not bias our inferences. The GLE events that can be simulated within the MAIRE tool are detailed in Table 2.3; we also provide an estimate of the maximum increase in count rate, as observed in the NMDB, as a reference.

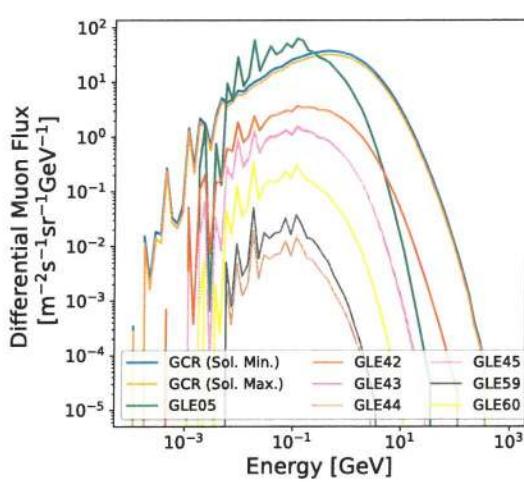
The GLEs incorporated in the MAIRE tool predate the existence of the HiSPARC network; as a result, the simulations are not directly informative on the GLEs we were investigating in Table 2.1. Nevertheless, the MAIRE simulations helped in our understanding of whether it was possible to observe a GLE using data

Table 2.3: The seven GLEs where MAIRE muon spectra were available, and the maximum observed increase in the neutron flux in the NMDB and the station where the increase was observed.

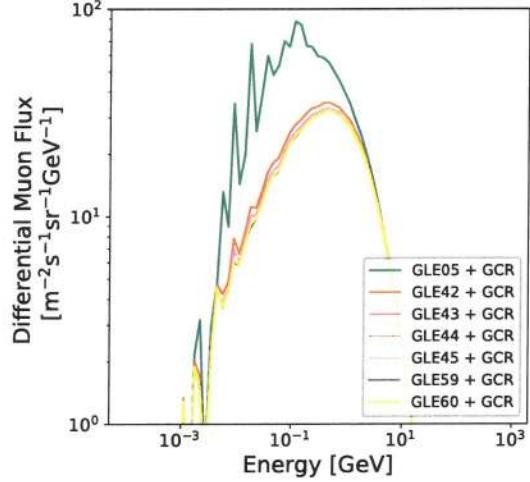
GLE	Date	Max. % change (station)
5	23/02/1956	~ 5100% (Leeds)
42	29/09/1989	~ 340% (Calgary)
43	19/10/1989	~ 90% (South Pole)
44	22/10/1989	~ 190% (McMurdo)
45	24/10/1989	~ 200% (South Pole)
59	14/07/2000	~ 60% (South Pole)
60	15/04/2001	~ 220% (South Pole)

acquired by the HiSPARC network.

We first used the MAIRE tool to simulate the spectra for the background GCRs and the seven GLEs at the Nikhef (501) HiSPARC station. Figure 2.28 shows the muon spectra for the GCR spectrum at solar minimum, and the additional muon spectrum for seven of the largest GLEs to date (which are additive to the GCR spectrum).



(a) Separate GCR and GLE Muon Spectra



(b) Sum of GCR and GLE Muon Spectra

Figure 2.28: Plots of the calculated MAIRE muon spectra for different incident GCR spectra. Blue and orange lines show muon spectra calculated for the incident GCR spectra during solar minimum and maximum, respectively. The other coloured lines show the computed muon spectra for the incident GLE spectra. (a) shows the individual muon spectra for GCRs and GLEs; (b) shows the combined muon spectra for the GCR at solar maximum and the GLEs.

The GCR-induced muon spectrum in Figure 2.28 roughly agrees with that computed using CORSIKA, which gave us confidence in the results of both simulations. We can see that the effect on the muon spectrum drastically varies for the seven GLEs, with GLE 5 clearly the largest event.

Prior to inferring any conclusions about the GLE-induced muon spectra at the HiSPARC stations, as the MAIRE tool provides the neutron flux, we verified the accuracy of the MAIRE results by comparing the observed increases in NM count rates to those predicted by MAIRE. To reduce any effects of strong anisotropy, when comparing to HiSPARC, we chose to analyse the NM count rate at a mid-latitude station, Kiel, which is located in Germany and is one of the nearest NMs to the HiSPARC network, at ~ 825 km away. The properties of the Kiel station are: $R_C=2.36$ GV, Altitude=54 m, Latitude=54.34°N, Longitude=10.12°E (NMDB, 2018); these are similar to the properties of the HiSPARC stations.

Table 2.4 shows a comparison between the observed neutron count increase, measured using data from the NMDB, and the increases predicted by MAIRE. The Kiel NM station was not online in 1956, during GLE 5, therefore this event has been omitted from these results. The MAIRE increases were calculated by comparing the integrated GLE flux to the integrated background GCR flux at that epoch, including the effects of the disturbances in the Earth's magnetic field from the planetary K-index (K_p) (Bartels et al., 1939). The muon count rate increases predicted by MAIRE are also shown in Table 2.4, for comparison.

We can see from Table 2.4 that there is a stark contrast between the observed NM increase and that predicted by MAIRE. One possible cause is inaccuracy in the predictions; however, this is unlikely as Lei et al. (2004) have shown the verification of their results versus observations, proving that MAIRE is robust. More likely, the difference is due to the anisotropy of the CRs initiating the GLEs, as MAIRE does not account for anisotropy in the simulations (Dyer et al., 2003; Lei et al., 2004). It is also possible that there is an unknown factor that causes this discrepancy, but we

Table 2.4: Observed and predicted increases in the CR count rates at the Kiel NM station, with rows ordered by the observed GLE magnitude. The observed neutron increases use data from the NMDB, where the errors are the measurement uncertainty from the 5-minute averaged data. The predicted data for the neutrons and muons are from the MAIRE simulations. The ratio column provides a conversion factor from the MAIRE predicted neutron increases compared to those observed. The fraction column shows the proportion of the MAIRE predicted muon increase compared to the MAIRE predicted neutron increase.

GLE	Neutron Increase			Muons Increase	
	Observed	Predicted	Ratio	Predicted	Fraction
42	$\sim 160 \pm 5\%$	88.7%	1.80	3.98%	4.45%
45	$\sim 45 \pm 2\%$	67.9%	0.66	0.817%	1.20%
60	$\sim 24 \pm 2\%$	18.8%	1.28	0.0814%	0.43%
43	$\sim 19 \pm 1\%$	59.6%	0.32	0.790%	1.32%
59	$\sim 8 \pm 1\%$	6.1%	1.32	0.00906%	0.15%
44	$\sim 6 \pm 1\%$	2.6%	2.30	0.00376%	0.14%
5	-	-	-	-	-

cannot postulate on this.

The ratio column gives a factor to convert from the MAIRE predicted increase to the NMDB observed increase, assuming that there are no other physical reasons why the predictions and observations differ. We assumed this could be used as a calibration factor, to ensure agreement between the MAIRE predictions and the observed data.

The ranked ordering of the events in Table 2.4, from strongest to weakest GLEs, shows the order of observed Kiel data and MAIRE-predicted data are in good agreement, suggesting that despite an incorrect magnitude of the increase, there is agreement between the two data sets. This supports the use of the calibration factor.

In addition, comparing the fraction of the predicted muon increase to the predicted neutron increase, shown in Table 2.4, we see the significantly smaller prediction for the muons. We know this is due to the energy spectrum of the PCRs, but the values in Table 2.4 provide a quantitative estimate of this and serves as further evidence to show that we may be unable to observe weaker GLEs with the HiSPARC network. However, Lovell et al. (1998) reported that during GLE 42,

several ground-based MDs across Canada and Russia measured increases of up to $\sim 30\%$ prior to the peak measured using NMs. Lovell et al. (1998) also showed that at the time of the NM peak, MDs measured a variation of magnitude of $\sim 8\%$. This also suggests a large underestimate in the MAIRE simulations. We see similar, low-magnitude predictions for the HiSPARC stations, which are shown in Table 2.5. There is good agreement between the predicted muon increase at the Kiel and HiSPARC stations. However, this agreement seems to depend on size of increase, with the weaker events showing less good relative agreement; this is down to the error in the predicted increase for such low-magnitude GLEs.

Table 2.5: The MAIRE predicted increase in the muon flux at Kiel and two HiSPARC stations. The rows are ordered by the observed NM magnitude from Table 2.4.

GLE	Predicted Muon Increase		
	Kiel	HS 501	HS 14001
42	3.98%	3.97%	4.06%
45	0.817%	0.798%	0.847%
60	0.0814%	0.0768%	0.0808%
43	0.790%	0.774%	0.817%
59	0.00906%	0.00563%	0.00683%
44	0.00376%	0.00241%	0.00305%

As the ratio column in Table 2.4 provided us with a calibration to recover the observed increase in neutron counts from those predicted in the MAIRE simulations, we used the assumption that the calibration factor may also be applied to the predicted increase in the muon counts, to recover the ‘true’ values. This assumed that both detectors only observe muons from Air Showers (ASs) (i.e. from the ‘sky’) and we do not observe regenerative muons.

In particular, we also intended to apply this calibration to the predicted magnitudes for the HiSPARC stations. To test that this was acceptable, and verify that the HiSPARC stations are sufficiently close to Kiel such that the deviation in the prediction is minimal, we investigated the dependence of the GLE magnitude on the latitude, longitude, and altitude of the location of the detector. The re-

sults of the simulations are shown in Figure 2.29. This plot shows the predicted increase in muons during GLE 42 when varying geographical parameters in the simulations. The blue square, orange up-triangle, and green down-triangle show the predicted magnitude at the location of the Kiel NM station, and HiSPARC stations 501 and 14001, respectively. The black circles represent the predicted magnitude when changing a single property (either latitude, longitude, or altitude) of the Kiel station during the simulations. The top panel in Figure 2.29 shows the relationship with latitude, the middle panel shows the relationship with longitude, and the bottom panel shows the relationship with altitude.

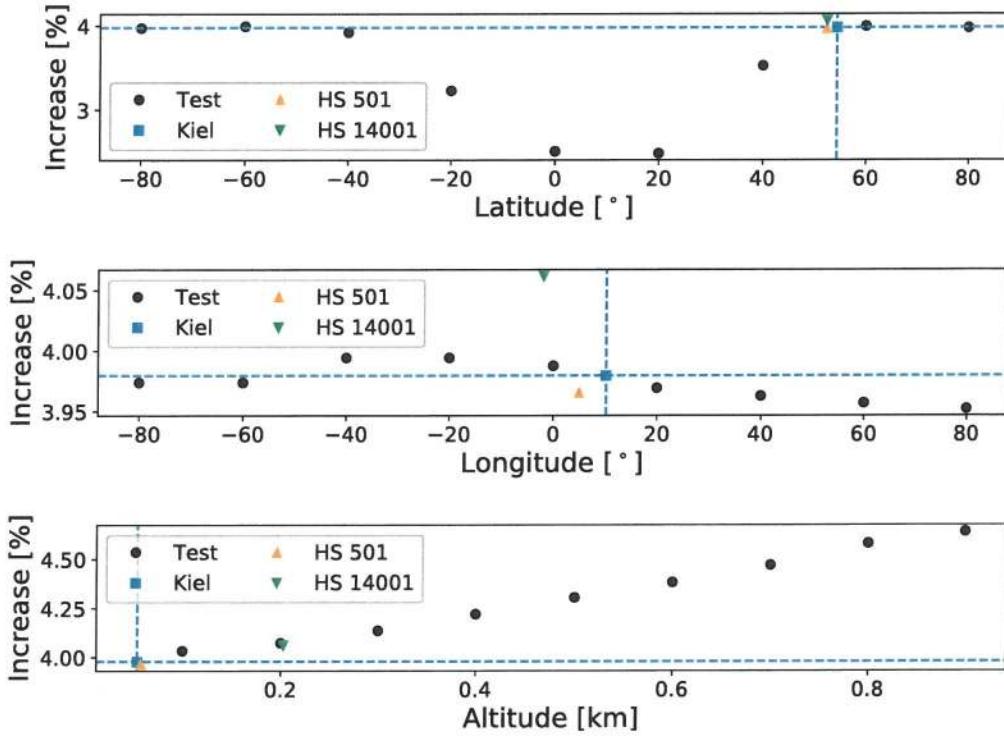


Figure 2.29: Shows the relationship between the predicted magnitude of GLE 42 and geographical location. Blue square shows the predicted muon increase at the Kiel station, orange up-triangle for HiSPARC station 501 (Nikhef), and green down-triangle for HiSPARC stations 14001 (University of Birmingham). The black points show the results when varying one property of the Kiel simulations, in the top panel we vary latitude; middle panel the variation with longitude; the bottom panel shows the variation with altitude.

From Figure 2.29, we found that the largest change in the predicted GLE magnitude occurred as a result of varying the latitude, which is expected due to the change in rigidity cut-off. The variation with longitude was marginal because MAIRE does

not account for anisotropy and the variation with altitude was notable. However, the important result from Figure 2.29 was the good agreement in the predicted muon increase between Kiel and the HiSPARC stations. The biggest deviation was for the station in Birmingham, which was solely due to the altitude difference, according to the lower panel in Figure 2.29. The results show that there is a lack of variance in the model predictions over the geographical scale between all three locations, hence we were confident enough to apply the same neutron-based calibration factor to all the muon predictions.

Based on the results in Figure 2.29 we used the neutron-based calibration factor to correct the muon predictions from MAIRE for the muon count rate at Kiel and the two HiSPARC stations, and the calibrated values are given in Table 2.6. GLE 5 has been omitted from these results as the Kiel station was not online then, so we cannot calibrate the MAIRE predictions. These values effectively represent a zeroth- or first-order approach to estimating the muon increase that would be observed for these GLEs with the data and information we have available. The calculations relied on the assumption that the application of the calibration factor was acceptable, and that anisotropy does not have a significant effect on the increase observed between the Kiel NM station and the HiSPARC stations in the Netherlands and the UK. In addition, we cannot quantify the uncertainty in the MAIRE simulations, but assume that the uncertainty is small.

The calibrated predictions for GLE 42 are in good agreement with the observed magnitudes using MDs, at the time of the GLE peak measured with NMs, of \sim 8% (Lovell et al., 1998). However, the values do not agree with the larger magnitude, earlier peak observed by MDs of \sim 30% (Lovell et al., 1998). This highlights a limitation in the MAIRE simulations for the muons, but demonstrates that MAIRE provides a good prediction of the magnitude at the time of the peak observed by NMs and gives a conservative estimate of the MD peak; being optimistic, we might expect to observe larger magnitudes than these predicted values with HiSPARC.

Table 2.6: The MAIRE predicted increase in the muon flux, adjusted for the calibration factor between observed and predicted neutron increases at the Kiel station, in Table 2.4. The rows are ordered by the observed NM magnitude from Table 2.4.

GLE	Calibrated Muon Increase		
	Kiel	HS 501	HS 14001
42	7.18%	7.15%	7.33%
45	1.47%	1.44%	1.53%
60	0.147%	0.139%	0.146%
43	1.42%	1.40%	1.47%
59	0.0163%	0.0102%	0.0123%
44	0.00678%	0.00435%	0.00550%

The effect of the GLEs shows only a small increase in the muon count rate, and for most of the events induces an increase of $< 2\%$ in the ground-level muon flux. Furthermore, for the weaker, more frequent standard of GLEs an increase of $\ll 1\%$ was predicted. The exception in these events is GLE 42. We expect that we would have seen the increase in the HiSPARC data for GLE 42, but it pre-dates the HiSPARC project and thus it cannot be verified. We expect the same would have been true also for GLE 5, but we do not have the Kiel data to form the calibration, to predict the increase. GLE 5 was an exceptionally large event, for which we have not seen anything similar in over half a decade. Such events are rare, and we expect that the energies involved would have shown a significant increase in the HiSPARC count rate.

These simulations, combined with the values given in Table 2.1, suggest that we would have expected an increase in the muon spectrum of $< 0.1\%$ for both GLE 71 and 72, which ~~rules~~^{makes} their observation with HiSPARC ~~as~~ extremely unlikely and helps explain why we were unable to observe them in the preceding investigation in this chapter. GLE 70 induced a larger increase in the Kiel neutron counts on the order of $\sim 30\%$, but comparing this with similar values in Table 2.4 and Table 2.6 suggests we would still only expect an increase in the HiSPARC count rate between $\sim 0.1\text{--}2\%$. In each of these cases it is very unlikely that we would have observed the GLEs using the HiSPARC network.

From these results we must therefore conclude that the HiSPARC network, in its current state, is unable to observe the more abundant standard of GLEs, but it is likely able to observe the rare, highly energetic events, such as that seen in 1956 leading to GLE 5.

2.9 Conclusion

We have presented a feasibility study on using the existing HiSPARC network of muon detectors to monitor space weather events. This was achieved through calculating the observing properties of HiSPARC stations, investigating the presence of space weather signatures in the existing data for five HiSPARC stations, both in their raw form and after performing corrections for atmospheric effects, and by performing cosmic ray air shower simulations.

Using simulations of the interactions of CRs with the Earth's magnetosphere in the PLANETOCOSMICS tool, we were able to calculate the rigidity cut-off and AVDs of the HiSPARC stations. We showed that the rigidity cut-off limits the PCRs to particles with energies on the order of and above $\sim 10^9$ eV. The AVDs were useful in demonstrating that the observable, lower rigidity PCRs are deflected significantly by the Earth's magnetosphere and we may observe solar eruptive events when the station is not pointing in a direction in-line with the Sun.

In the raw HiSPARC data, we found that we were unable to clearly detect signatures of the FDs or GLEs that have occurred over the lifetime of the HiSPARC network. It was observed that a major obstructing factor was due to atmospheric effects causing additional variations in the data. Using a linear relationship between the logarithm of the normalised CR counts against both the zero-centred pressure and temperature, the atmospheric effects were corrected in the data.

After the correction of the atmospheric effects, when investigating the corrected HiSPARC data we found that we were still unable to observe any GLEs. The same was true for most of the FDs; however, we do speculate that we were able to observe

See my previous point about this

a signature of the FD in December 2014.

A further study was conducted to understand the flux of muons at ground level, and in particular how the flux varied during GLEs. Using Monte Carlo simulations of particle transport through the atmosphere for incident PCRs, with CORSIKA, we were able to interrogate the number and radial distribution of muons arriving at ground-level per PCR. This highlighted that the flux of muons is very low for PCRs with energies less than, or in the region of, the rigidity cut-off of the HiSPARC stations. This analysis therefore showed that any increase in the CR count from GLEs would likely be very low, due to the diffuse nature of the air showers induced by the low-energy PCRs that primarily make up the GLE-inducing SEP spectra. To observe events with HiSPARC we are excluded to only the most energetic events that are sufficient to produce large muons ASs.

Expanding on this analysis, a comparative study was conducted using the MAIRE tool. This allowed us to directly predict the increase in CR counts based on simulations of incident PCRs, comparing the output muon flux from GCRs and GLEs. A calibration was necessary to ensure that the MAIRE predictions were consistent with the NMDB observations; however, upon applying this calibration, we were able to show that the predicted increase in the muon count rate was significantly lower than the neutron count rate, for each of the six GLEs studied. We assumed that it was acceptable to apply the same calibration to the predicted values for two HiSPARC stations, which was validated with simulations across various latitudes, longitudes, and altitudes. This led us to conclude that the HiSPARC network is only capable of observing the most energetic GLEs, which occur less frequently than their lower energy equivalents.

We leave the reader with the following points:

1. The rigidity cut-off and AVDs of the HiSPARC stations were calculated using PCR transport simulations. We found the HiSPARC stations generally have a cut-off rigidity in the range, $R_C \sim 3.0\text{--}3.5$ GV, setting a limit on the PCRs

observable on the order of $\sim 10^9$ eV. The asymptotic viewing directions for the HiSPARC stations are close to equatorial for low rigidity PCRs, between $\pm 20^\circ$ in latitude, and tend towards the station's zenith for rigidities greater than 20 GV.

2. Investigations of the raw and atmospheric corrected data showed no signatures of GLEs in the HiSPARC data. We propose that one FD was observed as a $\sim 2\%$ decrease in the atmospheric corrected data; however, none of the other events were observed in the HiSPARC data.
3. The CORSIKA air shower simulations showed that the flux of muons at ground level from low-energy CRs ($\sim 10^9$ eV) is very low, with diffuse air showers. With the current configuration of the HiSPARC network, which strongly relies on the triggering of multiple detectors within a station, the observations are biased to higher energies, hence showing a limitation of the HiSPARC network for space weather observations.
4. The MAIRE simulations showed that for some of the largest GLEs to-date, the predicted increase in the HiSPARC count rate was, on average, no more than $\sim 1\%$. Only the most energetic events, with a low occurrence, would induce an increase in the HiSPARC count rate by $> 5\%$. This showed that the HiSPARC stations are therefore generally incompatible with monitoring the lower limits of space weather activity, and can only detect the rarer, more extreme events.

This investigation has shown that the feasibility of using the HiSPARC network as a reliable tool for the monitoring of space weather is low; however, we have shown that for some specific cases, we expect to be able to make observations. One of the limiting factors for the application of HiSPARC for space weather detections is the current detector and trigger configuration. The positioning of the stations and the requirement of the time coincidence between events in multiple stations (2 or more)

biases the sensitivity towards higher energy PCRs. Since 2016, several HiSPARC stations have been providing the count rates of the individual detectors (the singles rates), which overcomes this bias; however, these data are often inconsistent between stations, have a worse signal/noise ratio, and show a strong diurnal variation due to thermal noise.

The current method of removing the temperature-induced diurnal variation uses the atmospheric temperature, and not the temperature inside of the roof-boxes, as it is the only temperature data available. Monitoring the temperature in the boxes themselves would provide a more accurate measure of the PMT temperature and hence the thermally induced noise, thus it will show a stronger relationship and provide a more accurate removal of this variation in the data. To overcome these effects, in the next chapter, we investigate the benefits of changing the configuration of a HiSPARC station, to determine whether we can improve the capabilities of the HiSPARC network to monitor space weather.

3 HiSPARC Station 14008

3.1 Introduction

It was shown in Chapter 2, using data acquired by the High School Project on Astrophysics and Research with Cosmics (HiSPARC) network, that in its original configuration, HiSPARC was not adequate for observing space weather events. In part, we showed that this was due to the low-magnitude of the increase in the expected muon flux during such events. Also relevant was the current configuration's bias towards higher energy Cosmic Rays (CRs) and the sensitivity of the detectors to variations in meteorological conditions.

To some extent, it was possible to eliminate the variation in CRs due to meteorological variations in the HiSPARC data; however, it was shown to not always be effective, as different detectors in the HiSPARC network displayed different responses to pressure and temperature variation and the correlation between atmospheric temperature and CR count was weaker than the counterpart for pressure.

Thermal fluctuations in the atmosphere induce thermal noise in the Photo Multiplier Tubes (PMTs) and although the temperature inside the HiSPARC roof boxes has not been measured, it is suspected that the PMTs can get quite hot, in particular when the roof boxes are in direct sunlight. We reported in Chapter 2 that the singles data represented our best possibility of observing lower-energy Primary Cosmic Rays (PCRs); however, these data are most susceptible to the induced thermal noise as the temperature of the PMT changes. Without measuring this temperature directly, a complete correction of this effect in the existing HiSPARC data was not

possible.

In this chapter we describe an alternative configuration of HiSPARC station, which was devised and tested, to minimise these limiting effects. An instance of thermal noise in a single PMT will be random, and uncorrelated with an instance of thermal noise in another PMT. To exploit this, we stacked two detectors on top of each other and put them in coincidence to measure a single muon which traverses both scintillators, hence inducing signals in both PMTs.

3.2 Aims

The principal aim of building a new HiSPARC station was to investigate whether an alternative configuration of a HiSPARC station could minimise atmospheric variations in the data. In addition, we aimed to demonstrate a configuration that allowed for the observation of space weather events.

We aimed to set up a new detector, perform the relevant atmospheric corrections, where necessary, and review the noise properties of the detector. Furthermore, we also aimed to perform simulations of Ground Level Enhancements (GLEs) of varying physical properties to understand what magnitude of GLE could be observed with the new set-up. This would help us to understand how likely it was to observe the any space weather events with the alternative HiSPARC station configuration.

3.3 HiSPARC Station 14008 Set-Up

3.3.1 Configuration

The new configuration, of HiSPARC station 14008, is shown in Figure 3.1; it is composed of two detectors stacked on top of each other, both inside one roof box, and the signals from the PMTs are put in coincidence. This configuration is advantageous over the single-scintillator, single-PMT HiSPARC set-up, as it allows the measurement of single muons that traverse both scintillators.

We showed in Chapter 2 that the existing HiSPARC design, requiring coincident

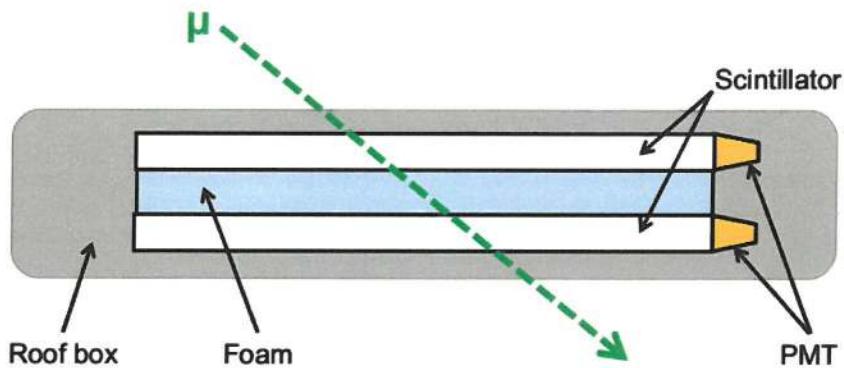


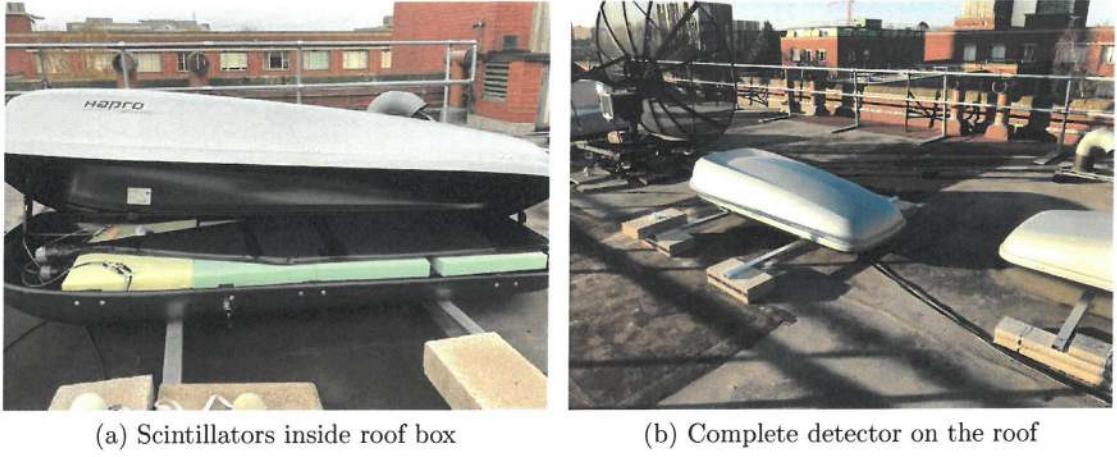
Figure 3.1: Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box.

triggers between detectors spaced tens of metres apart, biased the observations to higher PCR energies and single muons from lower energy PCRs could not be counted. Previously, we could only count single muons in the singles rates, but we have shown that the data were inconsistent between stations and it was difficult to disentangle the effects of temperature, PMT noise, and the diurnal effect in the data. This stacked-configuration design reduces the energy bias in the events data, as it no longer requires the large footprint Air Showers (ASs) to trigger multiple detectors, and provides a signal with fewer sources of noise than the singles rates, as it relies on the coincidence of two PMTs therefore minimising thermal fluctuations.

To protect the scintillators and PMTs, we sandwiched a layer of high density ($\rho = 38 - 40 \text{ kg m}^{-3}$, eFoam (2017)) foam, of thickness $\Delta x = 50 \text{ mm}$, between the scintillators, as can be seen in Figure 3.1 and Figure 3.2a. Upon the completed assembly of the detectors, they were placed within the roof box on the roof of the Poynting Physics building on the University of Birmingham campus (see Fig. 3.2).

Propagating charged particles lose energy in matter. Derived from the Bethe-Bloch formula (Bethe, 1932; Ziegler, 1999); we can estimate the amount of energy lost by a particle in a material as:

$$\Delta E = \Delta x S \rho \cos(\theta), \quad (3.1)$$



(a) Scintillators inside roof box

(b) Complete detector on the roof

Figure 3.2: HiSPARC 14008 assembly and configuration. (a) Shows the stacked arrangement of the scintillators within the roof box, between layers of protective foam. (b) Shows the complete detector on the roof of the Poynting building on the University of Birmingham campus.

where Δx is the thickness of the material, S is the stopping power of the material, ρ is the density of the material, and θ is the angle the particle travels through the material from the perpendicular direction.

Each of the plastic scintillators has a thickness of $\Delta x = 2.0$ cm, and density, $\rho = 1.03 \text{ g cm}^{-3}$ (Montanus, 2017). The stopping power of the scintillator for a minimum ionising particle is $S \sim 2 \text{ MeV g}^{-1} \text{ cm}^2$ (Fokkema, 2012; Montanus, 2017). The energy loss of a vertically incident muon in a single detector is therefore $\Delta E \sim 4 \text{ MeV}$. van Dam et al. (2020b) states the most probable energy loss of a vertically incident muon in a single scintillator is approximately 3.51 MeV.

Assuming a similar stopping power as above for the foam (Groom et al., 2001; Montanus, 2017), the muons will lose an additional ~ 0.4 MeV. In the complete configuration, as a muon traverses two scintillators and the foam, the estimated lower limit on the energy loss by muons in the detector is ~ 7.4 MeV. This new lower limit does not significantly change the values of the predicted GLE magnitudes in Section 2.8.4.

The standard HiSPARC station set-up is such that the PMTs are connected to the HiSPARC electronics box for data acquisition. In the standard HiSPARC station

configuration, the trigger rate of events is ~ 1 Hz. In this stacked configuration the trigger rate is significantly higher, ~ 80 Hz; hence the data produced is the equivalent of approximately half of the existing HiSPARC network. The HiSPARC servers could not cope with such a large quantity of data, therefore we had to reduce the data acquired by the HiSPARC box; however, we did not want to lose the original count rate of the stacked detectors. To acquire the data in this set-up, we used a Nuclear Instrumentation Module (NIM) crate, as shown in Figure 3.3, and a Raspberry Pi was used to store the data, which is discussed in Section 3.3.4.

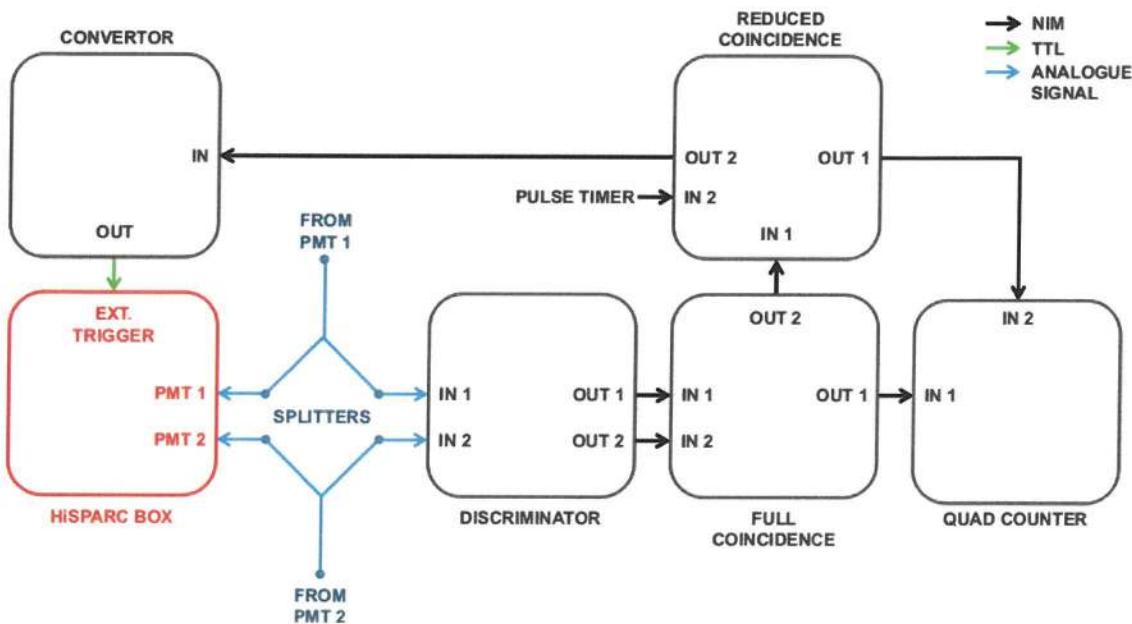


Figure 3.3: Schematic diagram of the HiSPARC 14008 station NIM crate configuration. Black-outlined boxes indicate the NIM crate modules, while red-outlined boxes depict HiSPARC hardware modules. Black arrows depict a NIM signal; green arrows show a TTL signal; blue arrows depict the analogue signal from the PMTs. The PMT signals are split and half of the signal interfaces directly with the HiSPARC electronics box and half the signal is passed through the NIM crate for processing.

The data acquisition is discussed in Section 3.3.4, but here we discuss the configuration of the NIM crate set-up. The analogue signals from the PMTs are split using a passive, equal-split resistive splitter such that half the signal is passed to the HiSPARC electronics box and half the signal is passed through the NIM crate. The analogue signal which is passed through the NIM crate goes first into a discriminator module (CAEN model N845) to only record signals that have an amplitude greater

than the trigger threshold. Due to the equal-balance resistive splitter the PMT signal is reduced in amplitude by a factor of 2; we used a discriminator threshold of -35 mV, i.e. half of the HiSPARC high threshold.

The discriminator outputs two NIM signals which are connected to the first coincidence module (LeCroy model 622). This records every coincidence between the two PMTs and the output from this module is directed to a NIM quad counter/timer module (ORTEC model 974). One channel of the NIM counter records the original coincidence counts from the first coincidence module.

A second terminal in the coincidence module was used to record a reduced count rate. This used a pulse timer (CAEN model 2255B) to create a gate signal with a duty cycle of $\sim 1\%$ (gate width = $45.0\ \mu\text{s}$; period = $4.86\ \text{ms}$). The coincidence between the original coincidence signal and the pulse timer gate ensures that the original count rate is reduced by a factor of ~ 100 . This was a sufficient reduction in the data for the HiSPARC servers to cope with. One output from this coincidence module is passed to the NIM counter, where it counts the reduced coincidences. The second output from the coincidence module is directed through a NIM-to-Transistor-Transistor Logic (TTL) converter and the output from this is used as an external trigger signal for the acquisition of data by the HiSPARC electronics box. This trigger was used to acquire the counts directly from the PMTs in the normal HiSPARC manner.

The use of these NIM modules introduces delays in the signal. Each of the NIM modules introduces a NIM-standard, typical input-output delay of $\sim 9.5\ \text{ns}$ (LeCroy, 1996; CAEN, 2011). In Table 3.1 we outline the delays that are introduced from the outputs of the PMTs to being registered at different end-points.

From Table 3.1 we see there is a delay of $\sim 36\ \text{ns}$ between the direct signal to the HiSPARC electronics box and the external trigger from the NIM crate. There is also an $\sim 8\ \text{ns}$ delay in between the full count and the reduced count. However, the delays introduced into the system are not actually a problem for counting muons

Table 3.1: Delays in the signals through different paths in the NIM set-up. The paths all start from the output of the PMTs, and are either direct or pass though the NIM crate before reaching their final end interface, therefore the path column is formatted as: start – direct/NIM path – end.

Path	Delay [ns]
PMT – direct – HiSPARC Box	16
PMT – NIM – Ext. Trigger HiSPARC Box	52
PMT – NIM – Counter (Reduced)	37
PMT – NIM – Counter (Full)	29

with the HiSPARC electronics box or the NIM counter. In the case of the HiSPARC electronics box the effect of the delays is mitigated by the low muon count rate and the wide pre- and post-trigger windows of the HiSPARC data acquisition software (discussed in Section 1.4.3). The HiSPARC data acquisition software uses pre-trigger ($1\ \mu\text{s}$), coincidence ($1.5\ \mu\text{s}$) and post-trigger ($3.5\ \mu\text{s}$) windows (Fokkema, 2012). This means that the signals coming from our NIM setup, with a maximum delay of less than 60 ns, will be easily captured within the $1\ \mu\text{s}$ pre-trigger window and thus counted. Using the NIM counter, as discussed in Section 3.3.4, we measure all counts in time intervals of 10-seconds, therefore a delay of ~ 8 ns does not impede our ability to count the events during the cadences.

3.3.2 Calibration

When setting up the HiSPARC station, it was required to set several operating parameters for the detectors and the HiSPARC electronics box. One such setting was the PMT operating High Voltage (HV). Each of the detector PMTs needs to be powered with a high enough operating voltage to provide an amplified signal, but not too high such as to over-amplify the noise.

In general, the PMTs have an advised operating voltage of around 700 V (Fokkema, 2019); however, best practise is to operate the PMT at the plateau region, whereby the counts/voltage no longer increases. As can be seen from Figure 3.4, neither of the PMTs have clear plateau regions, hence there was no obvious PMT set point.

The HiSPARC installation manual does, however, suggest to tune the PMT voltages such that the singles rates for each detector meet the following criteria: singles rate of 100–130 Hz for signal above the high trigger threshold, and singles rate of <400 Hz for signal above the low trigger threshold (Fokkema, 2019).

In order to calibrate the PMTs to the correct level, we measured the singles rates above the high and low thresholds as a function of PMT operating voltage, as is shown in Figure 3.4. The HV calibration plot shows the different performances one can get from different PMTs, therefore it was necessary to treat each PMT individually when calibrating.

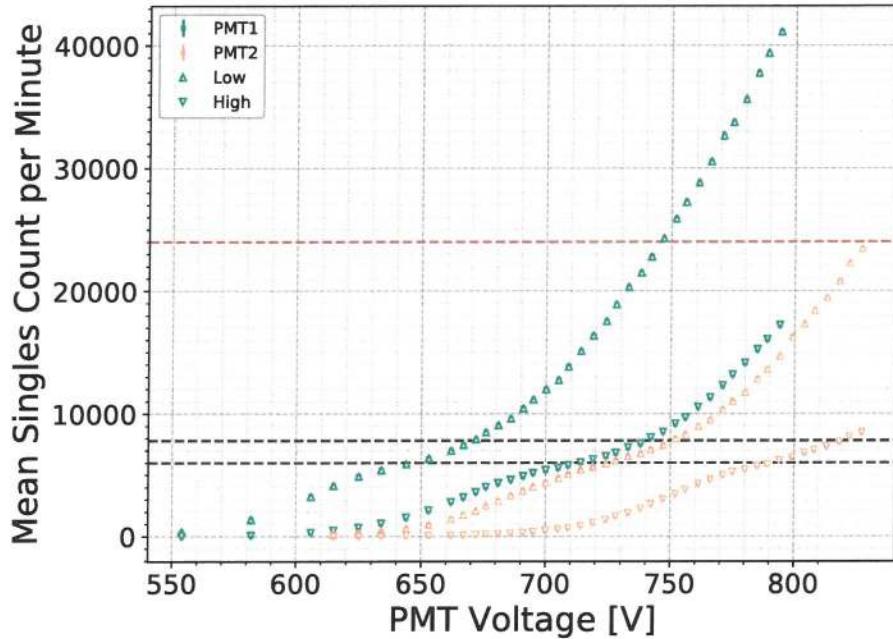


Figure 3.4: Voltage calibration curve for the PMTs of station 14008. The upper, red-dashed line indicates the upper limit for the low threshold singles rate (400 Hz), and the lower 2, black-dashed lines indicate the upper and lower bounds for the high threshold singles rate (100–130 Hz).

Initially the station was set-up supplying PMT1 and PMT2 \sim 725 V and \sim 790 V, respectively, based on the calibration in Figure 3.4. However, after some time the rates had drifted, perhaps due to early life-time variations in the PMT operations. After a re-calibration, since the end of 2019 the station has been consistently operating with PMT1 and PMT2 voltages of 725 V and 851 V, respectively.

how much time?

3.3.3 Monitoring Temperature

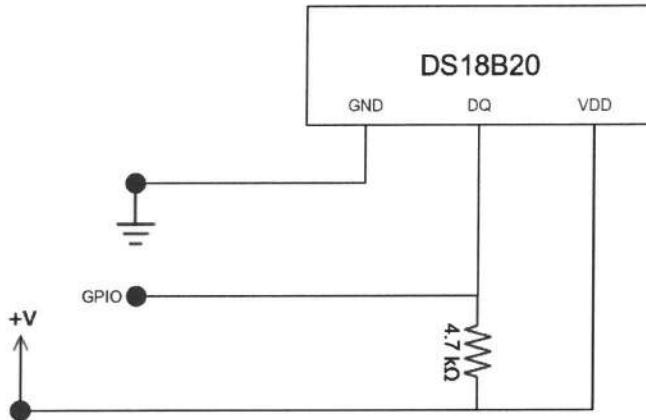
In Chapter 2, we suspected that the singles count rates were affected by the temperature of the PMT within the HiSPARC roof-boxes. Some of the existing HiSPARC stations monitored local atmospheric temperature, however none measured the temperature of the PMT inside the roof box. When building this new HiSPARC station, a temperature sensor was placed into the roof box which allowed us to monitor the temperature of the PMT more accurately.

Figure 3.5a shows the circuit diagram for the temperature sensor and Figure 3.5b shows the sensor inside the roof box. We used the DS18B20 temperature sensor with the one-wire telemetry protocol, which used a single wire to transmit the temperature readings to the microcontroller—a Raspberry Pi 4 (see Section 3.3.4). Three wires were used for the operation of the DS18B20: constant current voltage, ground, and data. The temperature was read on a 10-second cadence and recorded in degrees Celsius with a precision of 0.001° C.

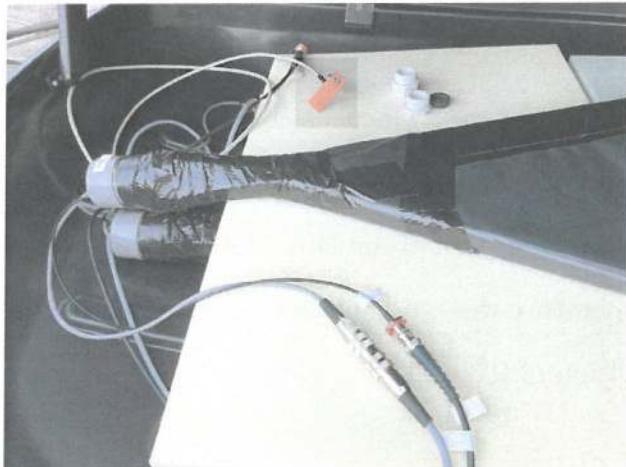
3.3.4 Data Acquisition

The new HiSPARC station uses two methods of data acquisition. The singles data and reduced coincidences data (events) are acquired using the typical HiSPARC data acquisition software, but the original coincidences, reduced coincidences, and the temperature data are all acquired by a Raspberry Pi 4. This was done as it allowed us to store the original coincidences data without overloading the HiSPARC servers. A schematic diagram showing the interfaces between the Raspberry Pi and the other hardware is shown in Figure 3.6.

The Raspberry Pi 4 was used to control the data acquisition by running continuous Python scripts; one for CR counts and another for the temperature data. The scripts configured the hardware and output the coincidences and temperature data to local files on the Raspberry Pi. The coincidences and temperature data are both recorded on a 10-second cadence.



(a) Circuit diagram



(b) Sensor within roof box

Figure 3.5: (a) Schematic diagram of the DS18B20 temperature sensor circuit, whereby the voltage drain (VDD), ground (GND), and data (DQ) pins connect directly to the voltage supply (+V), ground, and input/output (GPIO) pins of the Raspberry Pi board. (b) Shows the temperature sensor within the roof box, located by the PMTs. The temperature sensor is soldered into the circuit board seen in the top-middle of the image.

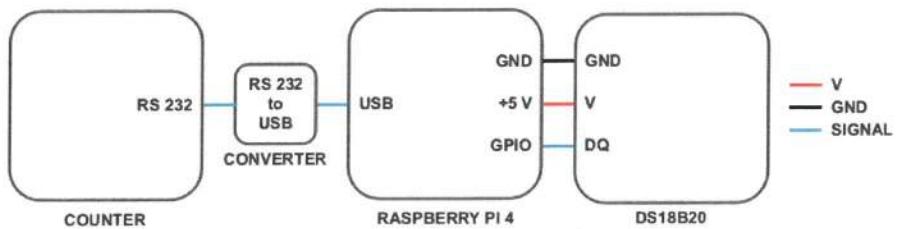


Figure 3.6: Schematic diagram of the HiSPARC station 14008 data acquisition interfaces. Red lines depict a 5 V signal; black lines show ground connection; blue lines depict the 1-wire protocol signal.

The Python scripts were written such that each new day generates a separate file for the coincidences data and temperature data. Within the coincidence files there are no headers and the data begins from line 1. The files contain four columns and the data stored in each column is listed in Table 3.2.

Table 3.2: Variables stored in the coincidences files of the HiSPARC 14008 instrument.

Column	Item	Unit	Type
0	Time Stamp	YYYY_MM_DD HH:MM:SS.fffff	String
1	Time*	Decisecond	Integer, eight digits, zero padded
2	Cumulative Reduced Count*	Counts	Integer, eight digits, zero padded
3	Cumulative Full Count*	Counts	Integer, eight digits, zero padded

* Since restart

The NIM counter records the cumulative coincidences count, therefore the reduced and full data stored are also cumulative and thus when reading the data, one must ensure that the difference is calculated between timestamps. In the event of hardware or software failure, or a reboot of the Raspberry Pi, when the Python script re-runs the NIM counter restarts all values from 0. When reading a file, one must ensure that checks are in place to handle any restarts from zero appropriately, such that no negative counts are calculated from one timestamp to the next during a restart.

Within the temperature files, there are also no headers and the data begins from line 1. The columns in the data files are outlined in Table 3.3.

Table 3.3: Variables stored in the temperature files of the HiSPARC 14008 instrument.

Column	Item	Unit	Type
0	Time Stamp	YYYY_MM_DD HH:MM:SS.fffff	String
1	Temperature	°C	Floating point

The coincidences and temperature data are stored locally, but they are also stored on the University of Birmingham Particle Physics servers as a back-up¹. Access to the data is not necessarily open-to-all, and to request access, one should contact the System Administrator for the Particle Physics Group Computing Facilities.

The reduced coincidences data are also acquired using the HiSPARC data acquisition software and are stored on the HiSPARC servers. The HiSPARC servers record this data as ‘events’ and the data can be accessed using the methods described in Section 2.3.

Data acquisition for station 14008 first began in March 2019. At this early stage of development, data were only acquired using the HiSPARC servers and not the NIM crate. The use of the NIM crate started in mid-September 2019, but it was not until mid-January 2020 that the temperature data were first acquired. The availability of data was interrupted during early months of 2020 due to the COVID-19 pandemic affecting our ability to perform crucial work on the station. It is therefore recommended to only use data after August 2020, when both coincidences and temperature data are regularly available.

3.3.5 Monitoring Pressure

As explained in the previous chapter, it was necessary to account for the barometric effect on the muon count rate, for both the singles data and coincidences data. To monitor the pressure, a nearby meteorological station was used, which is part of the Met Office Integrated Data Archive System (MIDAS) database, and acquired from the Science and Technology Facilities Council (STFC) and Natural Environment Research Council (NERC) Centre for Environmental Data Analysis (CEDA) archive.

The MIDAS station used is the nearest pressure monitor to the HiSPARC and provides a robust measure of the local atmospheric pressure. The station is located in Coleshill, Warwickshire (ID: 19187), nearby Birmingham International Airport, ~ 20 km from the HiSPARC detectors. The pressure is measured at the MIDAS

¹Disk location: /disk/moose/general/epesv001/datadisk/147.188.46.117_hisparc_pi/

There is a weather station on roof¹⁰⁷ of Gisbert Kapp
eee-Weather.bham.ac.uk/hix - it includes pressure. Would it have been

station level and a correction for altitude should be small.

The pressure data are recorded on a 1-hour cadence in units of hPa, with a precision of 0.1 hPa. The time variation of pressure is slow; hence, we linearly interpolated the data to provide a 1-minute sample.

3.4 Methodology

3.4.1 Atmospheric Corrections

~~After considerate review of the methods for temperature correction,~~ the method discussed and used in Chapter 2 was used here, i.e. using a linear relationship between CRs and temperature:

$$\ln \left(\frac{N}{N_0} \right) = -\alpha \Delta T, \quad (3.2)$$

where N is a single measurement; N_0 may be considered as the mean count rate over the given time-period of observations; $\Delta T = T - T_0$ is the deviation in temperature from the average (T_0) in the given time-period; and α is the temperature coefficient. However, the method was tweaked slightly. The steps for correcting for the effect of temperature were:

- To remove any long-term temperature trends over a month of data, the CR and temperature data were smoothed using a 24-hr moving mean. Using the linear relationship between the smoothed CRs and temperature (defined by equation (3.2), where the N and T are the smoothed data), the long-term temperature relationship was fitted and corrected in the data.
- After correcting for the long-term temperature relationship, for each day the temperature correction was applied again to remove the daily variations (i.e. using equation (3.2), where the N denotes the de-trended CR data and T denotes the raw temperature data; the de-trended CR and raw temperature

data were smoothed again using a 24-hr moving mean for N_0 and T_0). The temperature relationship was fitted and corrected in the data, day-by-day.

The above procedure was adopted because during the initial temperature corrections, it was found that without removing the long-term temperature relationship there was still some long-term covariance between the temperature and the CR count. In addition, we found that the previous correction procedure induced jumps in the data between days; this method of using smoothed data for the values of N_0 and T_0 ensured that there was a smooth transition in the corrected data between days.

Using the theory outlined in Section 2.6.2 we were able to perform the barometric correction of the singles data and coincidences data, whereby a linear fit was made using the model defined by equation (2.4). Similarly, the barometric correction was performed over durations of 1-month at a time.

For reasons discussed in Section 3.5 the temperature correction was, in practice, only applied to the singles data. When correcting the atmospheric effects on the singles data, the temperature correction was performed first and was followed by the pressure correction.

3.4.2 Observation Statistics

The probability distribution of the number of CR counts in a fixed interval of time follows the Poisson distribution, defined by:

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (3.3)$$

where k is the number of events, which is always an integer, and λ is the mean value of the number of events per interval, i.e. the expected number (Lista, 2016). Under the Poisson distribution, the mean and the variance are both equal to λ . For a large value of λ , the Poisson distribution can be approximated by a Gaussian distribution having mean, λ , and standard deviation, $\sqrt{\lambda}$ (Lista, 2016).

The Poisson distribution is also additive such that if two variables, n_1 and n_2 , follow Poisson distributions, with mean values λ_1 and λ_2 , respectively, then the sum also follows a Poisson distribution:

$$P(n; \lambda_1, \lambda_2) = P(n; \lambda_1 + \lambda_2), \quad (3.4)$$

where $n = n_1 + n_2$ (Lista, 2016).

Using this information, it was possible to use a Monte Carlo sampling algorithm to determine the mean level and noise of the HiSPARC 14008 station's data. We sampled the posterior using the PYMC3 No-U-Turn Sampler (NUTS) extension to a Hamiltonian Monte Carlo (HMC) sampling algorithm (Salvatier et al., 2016) with a Poisson distribution likelihood function. For the mean count rate we used a prior of $\mathcal{U}(3000, 6000)$ counts/min and for the noise rate we used a prior of $\mathcal{U}(0.0, 5.0)$ counts/min. In both instances, after a burn in of 2000 iterations, we used 5000 iterations on 4 chains to explore the posterior parameter space. Convergence was interrogated using the Gelman-Rubin \hat{R} diagnostic factor (Gelman & Rubin, 1992), using the criteria that chains did not converge if $\hat{R} > 1.01$.

In addition, with this knowledge artificial data were created, to simulate the detector's response to space weather events as a further study of the capabilities of the new station configuration. The artificial data were generated using the method discussed in Appendix B, to simulate the mean count rate and noise properties of the station. It was possible to inject GLEs into the artificial data with differing properties, to determine the likelihood of observing such events.

During the analysis of simulated data, we used a series of statistical tests to determine whether we could observe the injected GLEs. We used the fact that count data obey a Poisson distribution to compute the probability of statistically significant spikes in the data. The Poisson cumulative distribution function is given

by:

$$F(k; \lambda) = \sum_{i=0}^k \frac{\lambda^i e^{-\lambda}}{i!}. \quad (3.5)$$

Using this expression, the probability that a time interval observes k or more events, by chance, given the mean level, λ , is therefore given by:

$$p(k) = 1 - F(k-1; \lambda). \quad (3.6)$$

If $k \gg \lambda$ then $p(k) \rightarrow 0$. The probability that we fail to observe a cadence with k or more events is: $1 - p(k)$; thus the probability of failing to observe any time interval with k or more events in N -cadences is $[1 - p(k)]^N$. Therefore, the probability to find at least one event at or above k in N -cadences, by chance, is:

$$p_N = 1 - [1 - p(k)]^N, \quad (3.7)$$

where a low value for p_N indicates that the observed event is very unlikely to be a statistical fluctuation, and therefore a potential detection.

This can be generalised using the cumulative binomial distribution (Basu & Chaplin, 2017). The probability of finding at least r occurrences in N -cadences, by chance, at or above k , given the mean level, λ , is given by equation (3.8), which is equal to equation (3.7) when $r = 1$,

$$p[r; p(k), N] = \sum_{r=r}^N \binom{N}{r} p(k)^r [1 - p(k)]^{N-r}. \quad (3.8)$$

By applying equation (3.8) to the data, we were able to test whether there were any significant events against a certain probability threshold. Again, a low value for $p[r; p(\lambda), N]$ indicates that the event is very unlikely to happen by chance alone, and therefore a potential detection. The chosen probability threshold during all tests was 10%, i.e., this corresponds to a 10% chance of observing an excessive measurement as a statistical fluctuation.

The choice of N in this binomial test varies the required measurement threshold (Basu & Chaplin, 2017). In the tests performed we chose a value of $N = 720$ (i.e. ~ 2 hours for measurements with a 10-second cadence). This value of N was chosen as it is approximately the average Full Width at Half Maximum (FWHM) for GLEs (Strauss et al., 2017). In a 48-hour window, observing with a cadence of 10-seconds, we have $n = 17280$ total points; however, using $N = 720$ gives $n/N \approx 24$ independent time-windows to test. On average, we would therefore expect to measure approximately $0.1 \times 24 \simeq 2.4$ points exceeding the threshold due to statistical fluctuations. These represent unwanted, false-positive detections, but allow us to judge whether there is any potential significance in a given data set, depending on how the number of excessive measurements compares to this level. Furthermore, a space weather event, such as a GLE will produce excess events clustered in time; whilst false-alarm events will be randomly distributed.

Another frequentist test that was used makes use of the ~~assumption~~^{fact} that the Poisson distribution tends towards a Gaussian distribution when the mean value is sufficiently large. An excess in counts, compared to the mean value, can be quantified as:

$$s = k - \lambda, \quad (3.9)$$

where s is the excess in the signal, k is the measured signal and λ is the expected value, or background signal (Lista, 2016). The significance can then be approximated by:

$$Z = \frac{s}{\sigma}, \quad (3.10)$$

where σ is the expected standard deviation, which for a Poisson distribution is $\sqrt{\lambda}$. In this work we used both $Z = 3$ and $Z = 5$ significance levels to determine the existence of excess signals.

We also ran the statistics tests on 1-minute and 5-minutes averages of the artificial data. The statistics tests were run with the same underlying principles; however,

instead of using the Poisson cumulative distribution in equation (3.6), we instead used the Gaussian cumulative distribution for the averaged data with mean, $\mu = \lambda$, and standard deviation, $\sigma = \sqrt{\lambda/n}$, where n represents the number of data points used in the average.

3.5 Atmospheric Corrections

3.5.1 Temperature Correction

Using the method outlined in Section 3.4 we applied the temperature correction. The temperature correction was first applied between the coincidences data and the temperature measured inside the roof box. The typical relationship between the data and the temperature is shown for a single day in Figure 3.7.

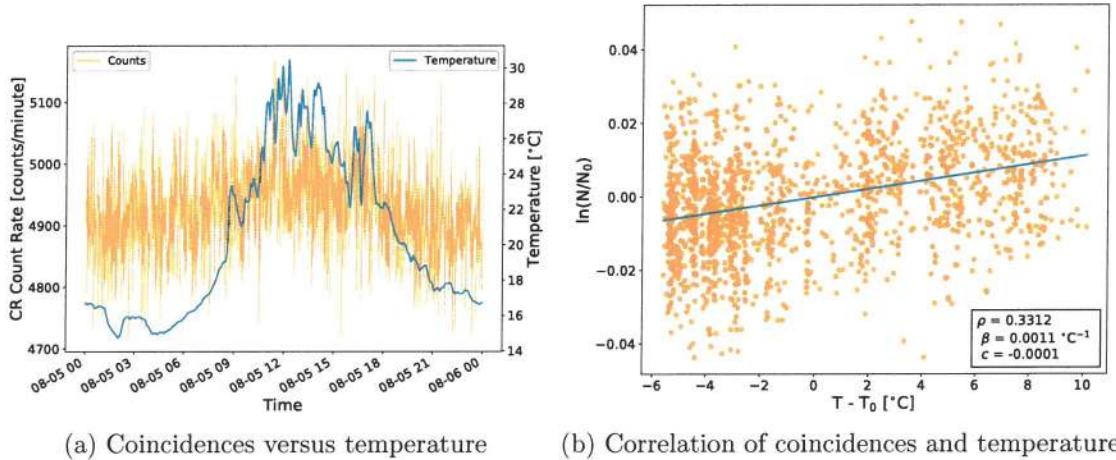


Figure 3.7: The relationship between the original coincidences data and the temperature within the roof box over a single day. (a) Shows the comparison between coincidences data (orange) and temperature data (blue), where the units of time on the x-axis are, MM-DD HH; (b) shows the correlation between the coincidences counts and temperature, and the fitted line to calculate the correction coefficient.

We see from Figure 3.7 that there is a weak correlation between the temperature within the roof box and the CR counts. Noting the common adage: “correlation does not necessarily imply causation”, we do not expect that the relationship between the coincidences and the temperature is causal. For this relationship to be causal, we require the increase in temperature of the PMTs to cause an increase in count rate

in the coincidences. We show later in Section 3.6.1 that the PMT thermal noise, which has a strong diurnal component, does not bleed through into the noise on the coincidences data. The weak correlation is a consequence of the rotation of the Earth meaning detectors look in different directions over the course of a day with a rise in temperature and CR counts at local noon (Parker, 1964; Mishra & Mishra, 2007, 2008). There is an increase in temperature around local noon as the Sun is overhead of the station, and the variation in the CR anisotropy in the interplanetary space causes a diurnal variation which is maximal when the detector is aligned with the Sun. We concluded that it was therefore not necessary to correct the coincidences data for the effects of temperature.

It was necessary to correct the singles data for the effects of temperature; this was one of the main reasons for introducing the measure of temperature within the roof box. One can see the relationship between the singles rates and the temperature in Figure 3.8.

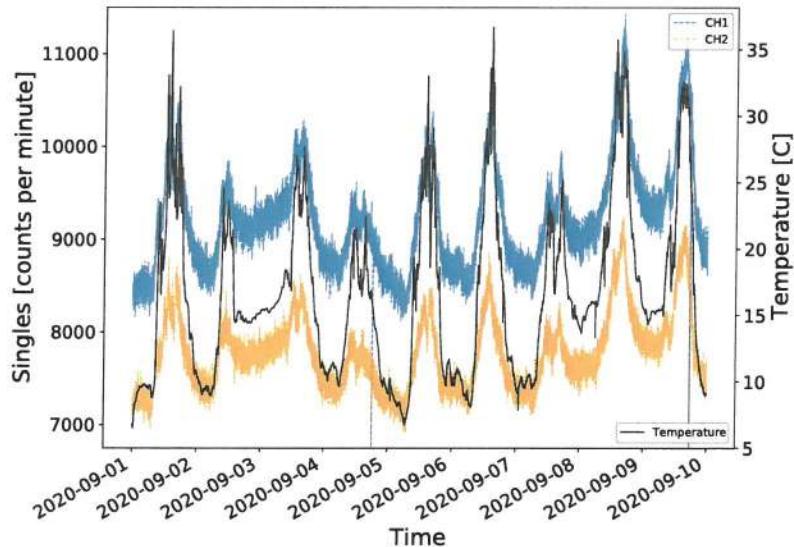
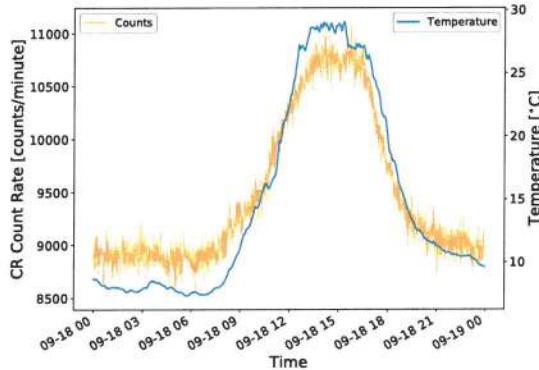


Figure 3.8: The relationship between the singles data (blue and orange lines) and the temperature within the roof box (black line). The units of time on the x-axis are, YYYY-MM-DD.

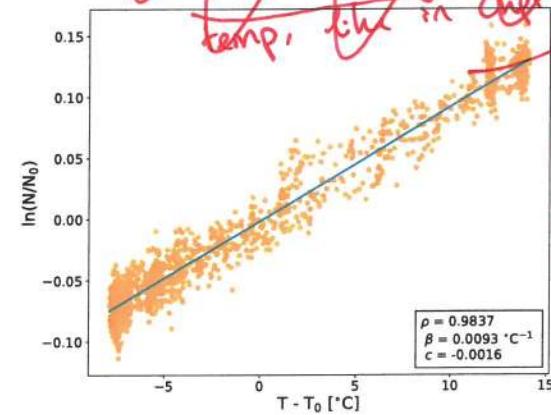
Figure 3.8 shows a strong relationship between the temperature inside the roof box (i.e. effectively the temperature of the PMTs) and the singles count rates. As

expected, the PMTs were sensitive to thermal variations, which induced thermal noise, and here we can see this is well-demonstrated. We showed in Chapter 2 that the atmospheric temperature was useful for correcting for the temperature variations in the singles rates, but not completely effective. The reason was because the temperature within the roof box is not the same as the atmospheric temperature.

The typical relationship between the singles data and temperature of the PMT is shown for a single day in Figure 3.9.



(a) Comparison of singles and temperature



(b) Correlation of singles and temperature

Figure 3.9: The relationship between the singles data and the temperature within the roof box over a single day. (a) Shows the comparison between singles data (orange) and temperature data (blue), where the units of time on the x-axis are: MM-DD HH; (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the correction coefficient.

The relationship between the singles data and the temperature inside the roof box is much stronger than the relationship between the atmospheric temperature and singles data, which was shown in Chapter 2. The temperature correction was applied using the linear fit between the singles data and the temperature inside the roof box. For comparison, and to show the success of this method at removing the temperature variation in the singles data, an example of the raw and corrected singles data are shown together in Figure 3.10.

In Figure 3.10 we see that the large, diurnal excursions are adequately removed from the singles data after this correction. This method of temperature correction

answered in
Viva

Could you also include here
the fit of using the 'external'
temp. like in ch. 2. Cover
to then see
the best
improvement in
this fit.
(as another
figure e.g.
3.9c).

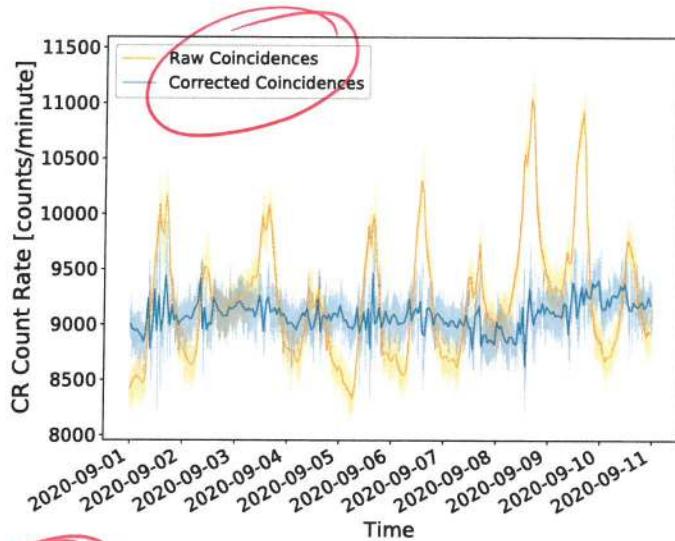


Figure 3.10: The singles data before (orange line) and after (blue line) the temperature correction process. The hourly resampled data are over-plotted to highlight the main variation in the data. The units of time on the x-axis are, YYYY-MM-DD.

was routinely applied to the singles data.

An additional benefit of the temperature monitor in the box of station 14008 was that it was also suitable for providing an estimate of the temperature inside the roof-boxes of the detectors that make up HiSPARC station 14001; hence the temperatures of those PMTs. Both station 14001 and station 14008 are located on the roof of Poynting physics building at University of Birmingham, therefore they are exposed to the same meteorological conditions and it is likely that the temperature inside one box is similar to the temperature inside each box. Figure 3.11 shows a comparison between the singles of the two stations, and the temperature within the box of station 14008.

In Figure 3.11 we see a good agreement between the singles acquired by both stations. Therefore, it is possible to also use this temperature data to correct for the effects of thermal fluctuations in the singles rates of HiSPARC station 14001.

3.5.2 Barometric Correction

Using the method outlined in Section 3.4 we were able to perform the barometric correction between the interpolated pressure data and coincidences and singles data.

'Can you add here (again) that the pressure data is from a nearby station - not from the box.'

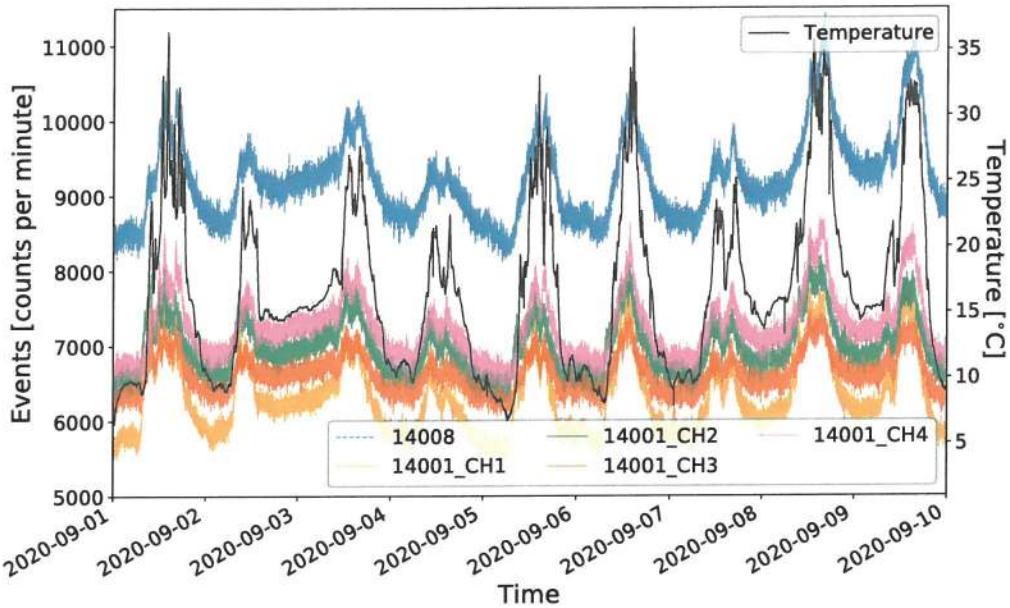
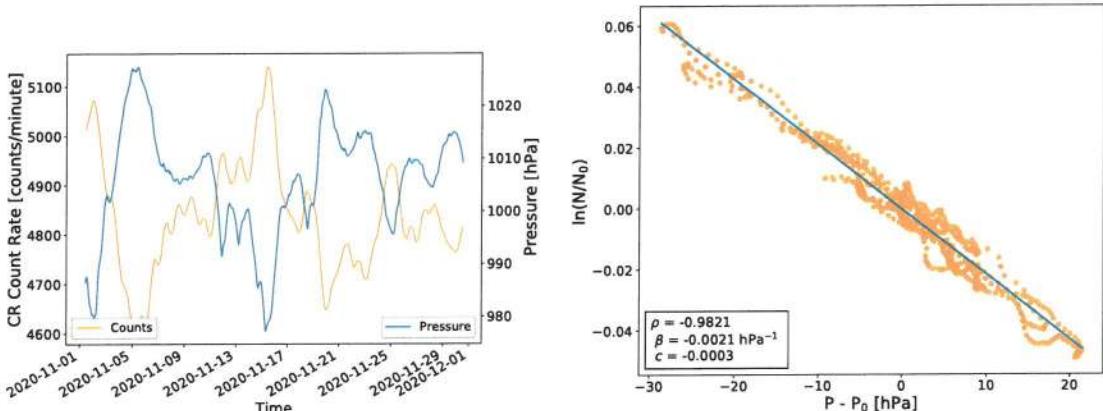


Figure 3.11: Comparison between HiSPARC station 14001 singles data and the singles data and temperature measured by station 14008. Black line: the temperature within the roof box of station 14008; dashed, blue line: HiSPARC station 14008 singles data; solid, orange, green, red, and purple lines: HiSPARC station 14001 singles data. The units of time on the x-axis are, YYYY-MM-DD.

Figure 3.12 shows a comparison plot of the smoothed coincidences and atmospheric pressure data sets.



(a) Comparison between coincidences and pressure
(b) Correlation between coincidences and pressure

Figure 3.12: The relationship between the smoothed, original coincidences data and the atmospheric pressure. (a) Shows the comparison between coincidences data (orange) and pressure data (blue), both with a 12-hour box-bar smoothing applied, to highlight the relationship and the units of time on the x-axis are: YYYY-MM-DD; (b) shows the correlation between the coincidences counts and pressure and the fitted line to calculate the correction coefficient.

As expected, Figure 3.12 shows the strong negative correlation between CR counts and atmospheric pressure. We were able to fit the linear model to the observed data, and the negative barometric coefficient was used to correct the data. For comparison, and to show the success of this method at removing the pressure variation, the raw and corrected coincidences data are shown in Figure 3.13. It is clear from Figure 3.13 that the large excursions are adequately removed from the data after the correction.

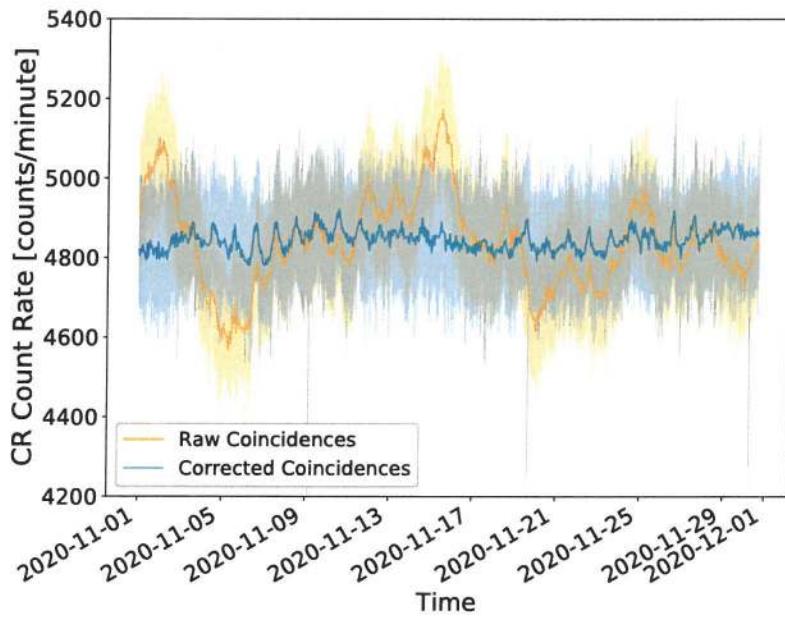


Figure 3.13: Showing the coincidences data before (orange line) and after (blue line) the barometric correction. The hourly resampled data are over-plotted to highlight the main variation in the data.

This method of barometric correction was routinely applied to coincidences data and singles data, to remove the barometric effect from the data acquired in this configuration.

3.6 Results

3.6.1 Observations

From the Cosmic Ray Simulations for Kascade (CORSIKA) air shower simulations performed in Chapter 2, we predicted an approximate ground level muon rate pass-

(as previous comment)

ing through a single HiSPARC detector of $\sim 85 \mu/\text{s}$ (for non-vertical, i.e. 70° acceptance cone simulations), and $160 \mu/\text{s}$ (for vertical simulations). These rates were comparable to the generally accepted, average ground level muon flux on the order of $\sim 70 \text{ m}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ (Cecchini & Sioli, 2000; Blackmore et al., 2015; Pereira et al., 2021; Particle Data Group et al., 2020).

In Figure 3.14 we show the corrected, original coincidence data which appears to have a mean count rate of $\sim 80 \mu/\text{s}$. In this plot we can see the diurnal effect. The diurnal effect measured here induced a variation in the CR count between $\sim 1\text{--}2\%$, which is larger than the $\sim 0.5\%$ diurnal variation, discussed in the literature (Mishra & Mishra, 2007, 2008; Dubey et al., 2016; Thomas et al., 2017), but is significantly lower than the variation observed in the standard HiSPARC events and singles data in Chapter 2. For any given epoch, the diurnal effect can be removed, if necessary, by subtracting a smoothed time series.

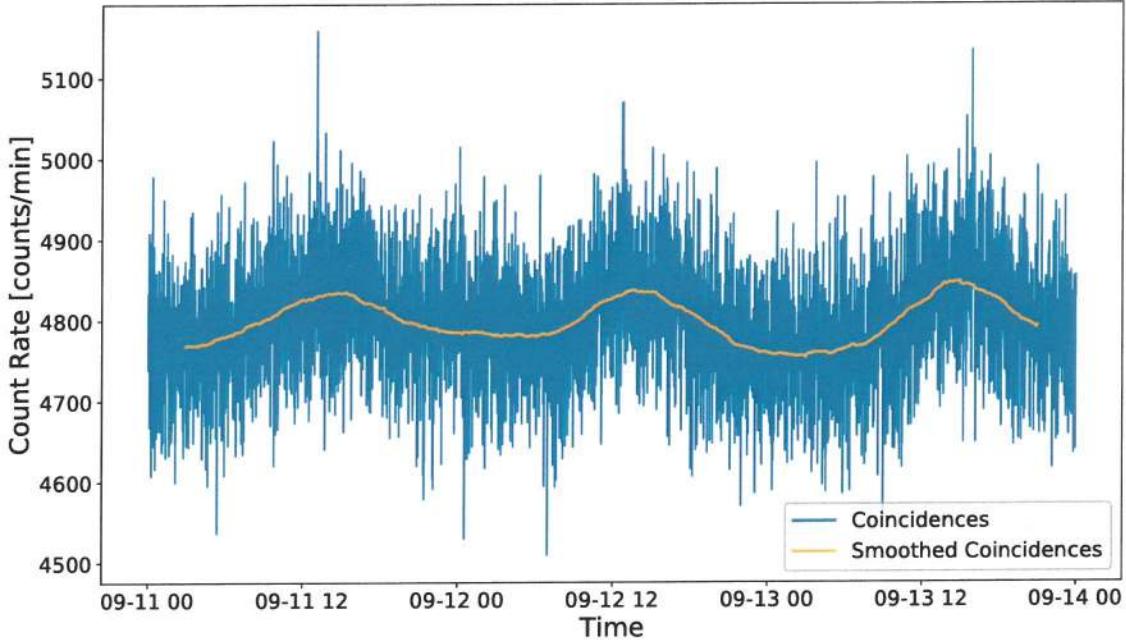


Figure 3.14: Time series of coincidences data, corrected for atmospheric pressure. The blue line shows the corrected data displaying the diurnal variation with peaks at around midday. The orange line shows the data smoothed using a 6-hour box-car. The units of time on the x-axis are, MM-DD HH.

As the counts follow a Poisson distribution we sampled the data using the `pymc3`

No U-Turn Sampler (NUTS) extension to a Hamiltonian Monte Carlo (HMC) sampling algorithm (Salvatier et al., 2016) with a Poisson distribution likelihood function. This allowed us to determine the mean count rate and convergence was interrogated using the Gelman-Rubin \hat{R} diagnostic factor (Gelman & Rubin, 1992) using the criteria that chains did not converge if $\hat{R} > 1.01$.

The median value of the posterior distribution for the mean value of the Poisson distribution of these coincidence data was 4797 ± 2 counts/min, where the uncertainties represent the 68 % credible intervals either side of the median. We therefore have a count rate of $\sim 80 \mu\text{s}^{-1}$ in this stacked detector configuration. This agrees remarkably well with the predicted value from the non-vertical simulations in Chapter 2, which represents a good approximation of the true muon flux at ground level. With a count rate of $\sim 80 \mu\text{s}^{-1}$, the Poisson noise is a rate of $\sim 9 \mu\text{s}^{-1}$, which represents $\sim 11\%$ of the signal.

These observations have used the original coincidences data, to determine the mean count rate. These data are stored only locally, but we also acquire the reduced count rates which are stored locally and separately on the HiSPARC servers. The reduced coincidences data sent to the HiSPARC servers use the NIM gate signal as a trigger which reduces the count rate by a factor of ~ 100 . The data stored locally were acquired slightly differently. As discussed in Section 3.3.4, the reduced counts (stored locally by the Python script) used the NIM counter to measure the rate of the external trigger signal (i.e. coincidences between the NIM gate signal, and the coincidences between the two PMTs). The HiSPARC events data use the trigger to read the events directly from the PMTs. Due to the delays in the signal in the NIM crate configuration, we investigated both data sets to ensure that they did not differ. In Figure 3.15 we show a comparison between the reduced coincidences data stored locally and those recorded as events data in the HiSPARC server.

the Python script doesn't actually store them. Maybe remove by the ...

The mean value of the Poisson distribution of reduced coincidences data is ~ 44 counts/min ($\sim 0.73 \mu\text{s}^{-1}$), which is a reduction by a factor of ~ 110 from the

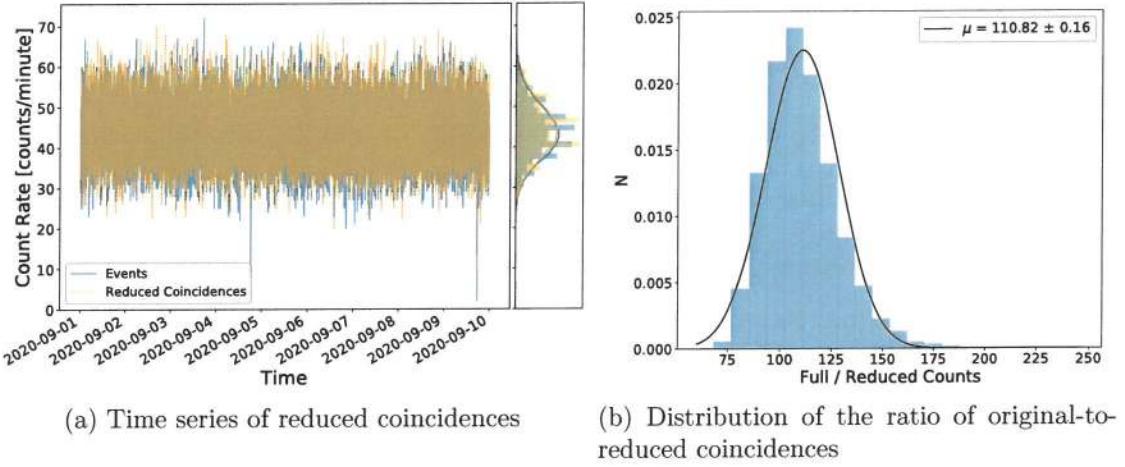


Figure 3.15: (a) Comparison of the reduced coincidences stored locally (orange) and as events data in the HiSPARC server (blue), where the units of time on the x-axis are, YYYY-MM-DD. (b) Distribution of the ratio of the original and reduced coincidence rates.

original coincidences data. We can see from Figure 3.15 that both the data stored locally (reduced coincidences) and the data sent to the HiSPARC server (events) are in agreement, with only the realisation of the noise being different between the data. The HiSPARC data acquisition software features a pre-trigger window (duration: $1\text{ }\mu\text{s}$), coincidence window (duration: $1.5\text{ }\mu\text{s}$), and a post-trigger window (duration: $3.5\text{ }\mu\text{s}$). The duration of these coincidence windows means that the small delays from the NIM modules does not affect the ability of the data acquisition software to record the data from the external trigger. This verifies that the locally stored reduced coincidences data and the events data stored in the HiSPARC server are measuring the same signal, with a count rate of count rate of $\sim 0.73\text{ }\mu\text{s}^{-1}$.

To understand the noise properties of this new configuration we investigated the random noise which is induced by random, spurious counts between both PMTs which do not coincide with the passage of a muon. This was achieved by adding a delay in the signal between the two PMTs, to ensure any coincident triggers were not due to true coincidences from the passage of a muon. By adding additional cables to the output from one PMT, a delay of $\sim 120\text{ ns}$ was added between the two signals. The FWHM of a typical pulse from the PMTs is $\sim 25\text{ ns}$, and the total duration

from beginning-to-end is on the order of 100 ns (van Dam et al., 2020b), therefore the \sim 120 ns delay was sufficient to remove true coincidences from the observations.

The delay was added between the two PMTs for around a week and the time series of the coincidences is shown in Figure 3.16. We can see that the noise is nominally \sim 1 count/minute.

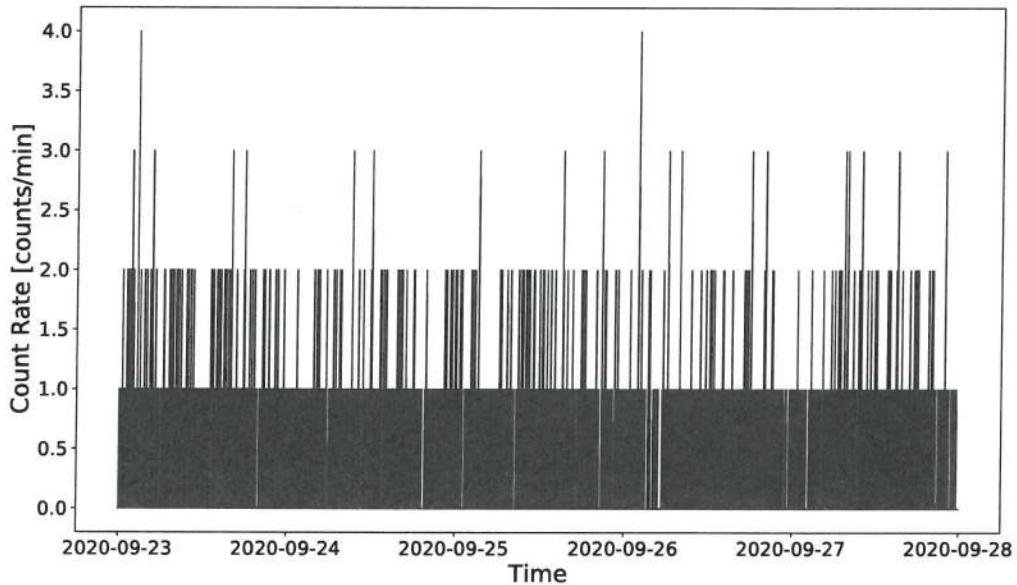


Figure 3.16: Time series of spurious coincidences data measured over five days. The units of time on the x-axis are, YYYY-MM-DD.

We know the noise must follow a Poisson distribution, therefore we aimed to quantify the mean value of the spurious noise. The noise was sampled to determine the mean of the Poisson distribution using the `pymc3` NUTS extension to a HMC sampling algorithm (Salvatier et al., 2016). Convergence was interrogated using the Gelman-Rubin \hat{R} diagnostic (Gelman & Rubin, 1992) factor using the criteria that chains did not converge if $\hat{R} > 1.01$. The distribution of the random coincidences is shown in Figure 3.17.

The median value of the sampled posterior of the mean value of the Poisson distribution of random coincidence is 0.259 ± 0.006 counts/min, where the uncertainties represent the 68% credible intervals either side of the median. This represents a low level of noise; under this Poisson likelihood function the probability of no noise is

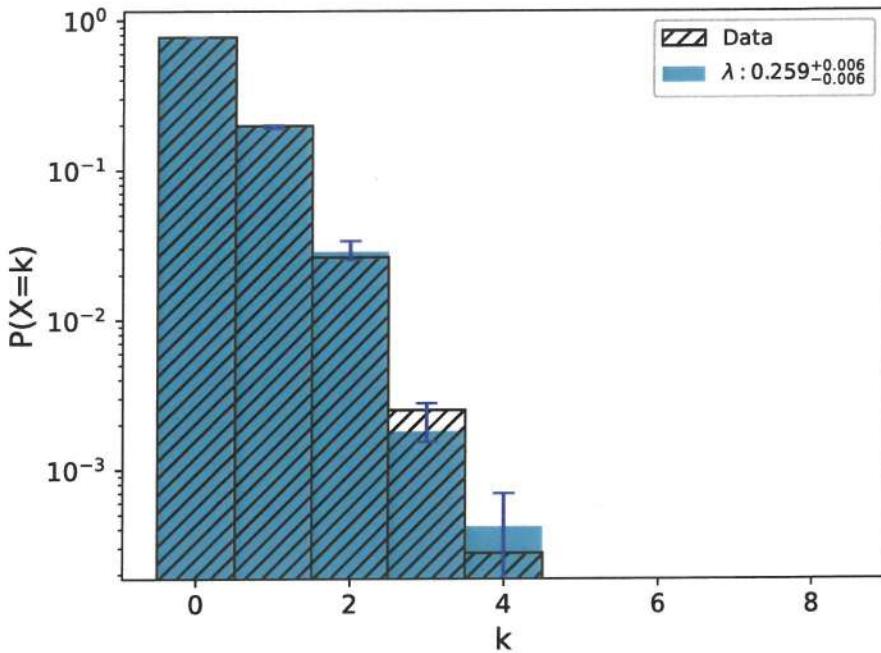


Figure 3.17: Distribution plot of the random coincidences data (hatched bars), and probability mass function of a Poisson distribution with mean equal to the median value of the posterior distribution after sampling (blue bars). Blue error bars represent the 68% credible intervals either side of the median.

$\sim 77\%$, 1 count/minute is $\sim 20\%$, and over 2 counts/minute is $< 3\%$.

It is also important to note that in Figure 3.16, there is no obvious diurnal pattern in the signal. This shows that as the PMT thermal noise increases around midday, which we see manifesting in the singles data, the increased thermal noise does not manifest in the spurious coincidences between both PMTs. This is important as it highlights that in this stacked detector configuration we have maximised our ability to observe single muons whilst reducing the effects of diurnal, thermally induced noise, which motivated not correcting for the weak correlation between the temperature of the PMTs and the coincidences data.

3.6.2 Comparison with Neutron Monitors

It is useful to compare the data from this new HiSPARC station to an existing Neutron Monitor (NM) detector in the Global Neutron Monitor Network (GNMN) (Mishev & Usoskin, 2020), as NMs are one of the most widely used sources of data

no GLE's. There were
other SpwX events!

for monitoring space weather. Unfortunately, there were no space weather events

from the beginning of HiSPARC station 14008 operation to the time of writing; however, we still show a comparison to a nearby NM station.

The Kiel NM station, in Germany, used in Chapter 2 had suffered difficulties with data consistency during this epoch, therefore another station was used in the analysis here. We chose to analyse the NM count rate at Dourbes, which is located in Belgium and is the nearest NM to the HiSPARC network, at \sim 525 km away from station 14008 in Birmingham. The properties of the Dourbes station are: $R_C=3.18$ GV, Altitude=225 m, Latitude=50.10°N, Longitude=4.59°E (NMDB, 2018). In Figure 3.18, a comparison is shown between the corrected HiSPARC coincidences data and the Dourbes NM station during November 2020.

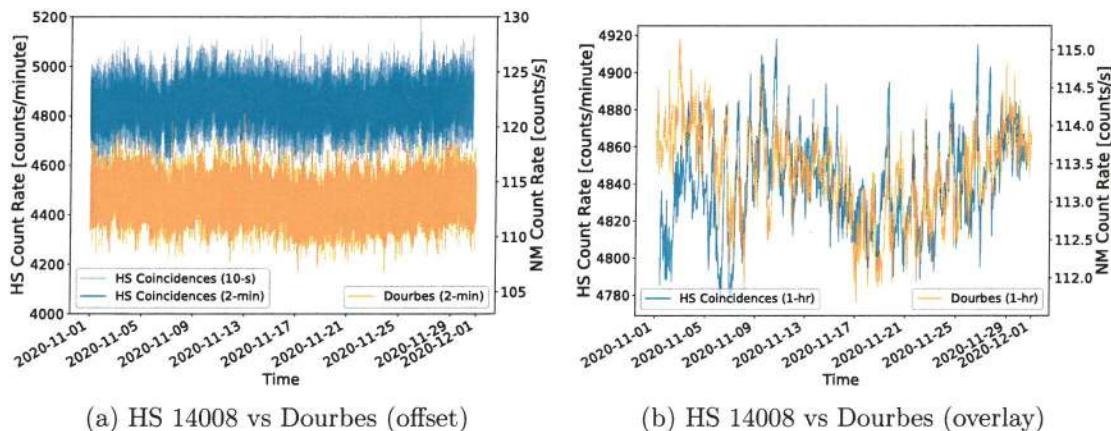


Figure 3.18: A comparison between the corrected HiSPARC station 14008 coincidences data (blue) and the pressure corrected neutron monitor data measured at the Dourbes NM station, Belgium. (a) Shows the two data sets offset and (b) overlayed hourly-averaged data, showing the similarities between the two data sets. Units of time on the x-axis are, YYYY-MM-DD.

The plots in Figure 3.18 show a good visual agreement between the two detectors. Despite a good visual agreement, the Pearson correlation coefficient was $\rho \sim 0.08$, highlighting that there was no correlation between the two stations, at the 2-minute level. When resampling the data to 1-hour and 1-day the correlation increased to $\rho \sim 0.40$ and $\rho \sim 0.41$, respectively. This showed a weak, low frequency correlation between the two stations. This weak correlation showed the two stations monitor

approximately the same background CR signal, which is relatively flat as the solar activity is low, therefore contributions from Solar Cosmic Rays (SCRs) are low. At the 2-minute level, the near-zero correlation demonstrates that the variations in the two signals are dominated by noise and there was no covarying signal.

This comparative analysis should be continually monitored, and particularly used as a reference when any space weather events are recorded with the GNMN. As it is the closest NM to the HiSPARC 14008 detector in Birmingham, it is useful to continue using the Dourbes NM station, but also to incorporate the use of data from the Kiel NM station, when issues with data quality are resolved. Near the maximum of Solar Cycle 25, expected 2023–2026, and likely to arrive in 2025 (McIntosh et al., 2020; Pesnell, 2020), we would expect the correlation to grow as the number of Solar Energetic Particles (SEPs) increases. It is therefore important that this configuration of HiSPARC detector is maintained until at least 2026, to ensure a complete study is performed when solar activity and space weather activity is high.

3.6.3 Single Station Space Weather Uses

Simulations of artificial data were performed to determine the magnitude of GLEs that may be observed in this new station configuration, as described in Section 3.4.2. We have shown the HiSPARC 14008 station has a background, mean count rate, $\lambda \sim 80 \mu/\text{s}$, and a noise of $\sigma \sim 0.26 \mu/\text{min}$ (i.e. $\sim 0.004 \mu/\text{s}$). These were used as inputs to the simulations, where GLEs were simulated with amplitudes of: 1.0%–5.0%, in intervals of 0.5%. In addition, we simulated GLEs with amplitudes of: 7.5% and 10.0%. The artificial data were created using the method in Appendix B and the statistics tests performed on the resultant data.

Running several iterations, it was possible to analyse the statistics for each amplitude of GLE, compared to the background signal/mean count rate of the detector, without a GLE. An example of the output from the statistical tests on a single iteration is shown in Figure 3.19, for a 10% GLE magnitude. Table 3.4 shows the average

number of measurements exceeding the various thresholds by chance for simulations without any injected GLEs.

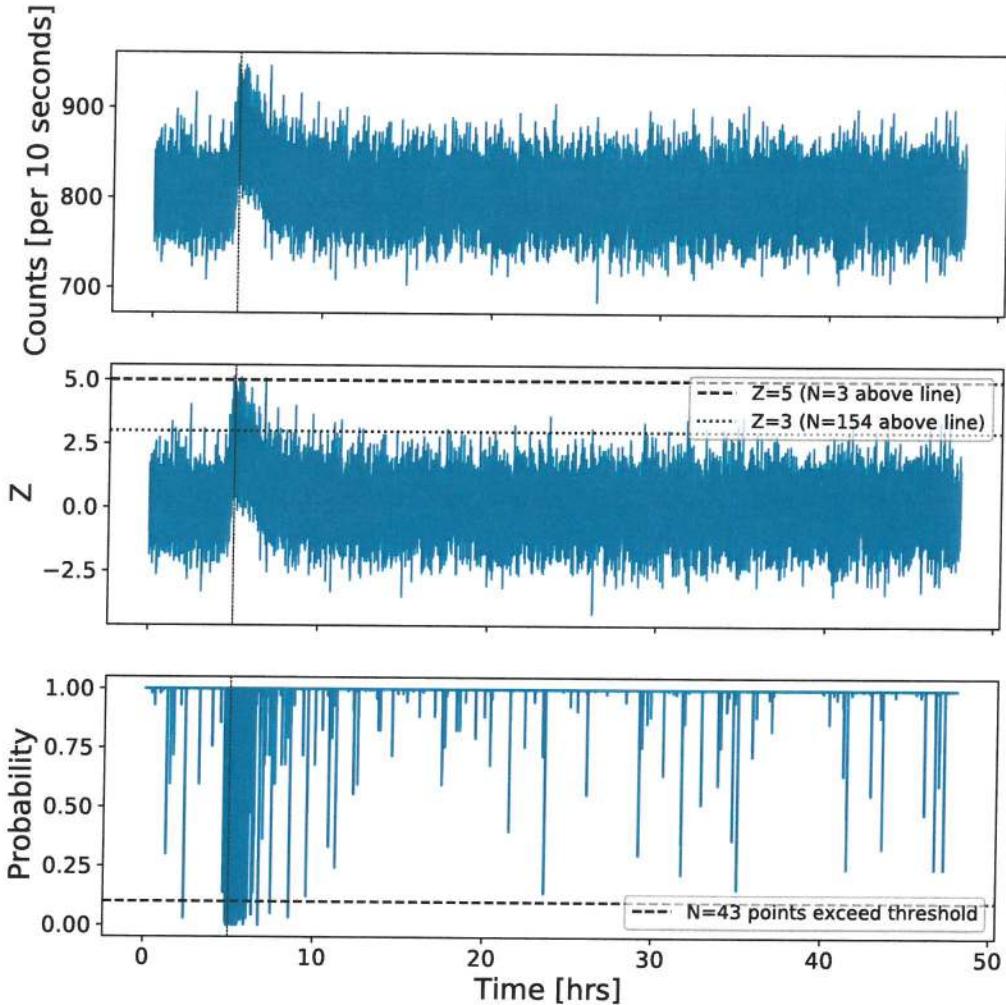


Figure 3.19: Single realisation of a simulation with a 10% GLE magnitude. Top panel shows the raw, artificial data from the simulation, with a 10-second cadence. Middle panel shows the number of measurements exceeding the $Z = 3$ and $Z = 5$ thresholds. The $Z = 3$ and $Z = 5$ thresholds are depicted as dotted and dashed lines, respectively. Bottom panel shows the number of points exceeding the $p = 10\%$ threshold in the binomial test, where a low probability value indicates a high chance the measurement is not statistical fluctuation. The dashed line shows the $p = 10\%$ threshold.

Table 3.4 shows what one sees (on average) by chance, with no GLEs present. In the case of the binomial test on the raw, 10-second cadence data, we found there was a 10% chance of observing 2^{+2}_{-1} or more measurements over the threshold, by chance, when using $N = 720$ points in equation (3.8) (i.e. ~ 2 hours). In Section 3.4.2, we

Table 3.4: Median number of excessive measurements in the artificial data with no injected GLEs. The values were acquired from 1000 iterations of simulated data, performing each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for the 10-s cadence data and averaging over 1-minute and 5-minutes. All uncertainties correspond to the 68% credible intervals either side of the median.

Data cadence	No. excessive measurements		
	Binomial	$Z = 3$	$Z = 5$
10-second	2^{+2}_{-1}	27^{+6}_{-5}	0 ± 0
1-minute	0^{+1}_{-0}	4 ± 2	0 ± 0
5-minute	0 ± 0	1 ± 1	0 ± 0

showed, on average, we would expect to measure approximately 2.4 points exceeding the threshold due to statistical fluctuations, in a 48-hour period; therefore the value in Table 3.4 agrees with the expected number. In the event of a GLE, we could claim a significant observation using the binomial threshold if we observed over ~ 12 measurements (i.e. as there is a probability of $\sim 8.5 \times 10^{-7}$ to observe random instances across the 24 independent realisations, which is comparable with the 5σ level). In particular, with the binomial test, we expect measurements of a GLE to give several excessive points close together, at the time of the GLE, and not scattered over the 48-hour period.

These results therefore allow us to judge whether there is any potential significance in a given data set, depending on how the number of measurements exceeding the thresholds compares to the values in Table 3.4. For example, in Figure 3.19, there were 43 points exceeding the 10% binomial limit, and 154 and 3 points above the $Z = 3$ and $Z = 5$ limits, respectively. This indicated the existence of significance in the data, in which a GLE was injected with a magnitude of 10%. On closer inspection one can see the grouping of the number of points exceeding each of the thresholds where the event occurred, at ~ 5 hours.

With artificial data, over a given two-day period, we can say that any epoch with a number of excessive points greater than these values can be treated as statistically of interest. We expect that any measurements which exceed the $Z = 5$

threshold should clearly be treated as a significant event claim, as within two days of artificial background data, we observed no random fluctuations in the noise exceeding this level. In addition, we see that there is a large difference in the number of measurements exceeding the $Z = 3$ threshold between the raw, 10-s data to 5-minute averaged data. We can be confident of claiming a statistically compatible or significant event if the number of excessive measurements exceeds 3- or 5σ .

For each GLE magnitude we ran 1000 iterations of the simulations and were able to average over the number of excessive measurements for a given GLE magnitude. In each simulation, the rise and decay times of the GLEs were randomly sampled (see Appendix B). This was done for the raw, 10-s cadence data and further simulations were performed for 1-minute averaging and 5-minute averaging. To summarise the results of all the simulations, Figure 3.20 shows the average number of cadences with excessive measurements against the simulated GLE magnitude for the 10-s cadence observations, 1-minute and 5-minute averages. The horizontal, dashed lines show the median number of points due to statistical fluctuations, i.e. without an injected GLE, and the horizontal, dotted lines represent the 68% credible intervals either side of the median. Each point shows the median number of excessive observations for different GLE magnitudes, also with error bars representing the 68% credible intervals either side of the median.

We see from Figure 3.20 that in the case where no averaging of data was performed, using the $Z = 3$ significance level, we begin to be able to differentiate the increase in the number of excessive measurements from simply statistical fluctuations for a GLE with magnitude of $> 3.0\text{--}3.5\%$. However, in the $Z = 3$ test we cannot claim to have observed a statistically significant event (i.e. to 5σ) until a GLE with magnitude of $\lesssim 4.0\text{--}5.0\%$. Using the binomial test, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $> 4.5\text{--}5.0\%$. To be able to see the results clearly, we only show the results for GLE magnitudes of up to 5 %, and in Figure 3.20a, we see using

I guess you are saying "approximately less than 5%, but greater than 4%"
Not very clear notation.

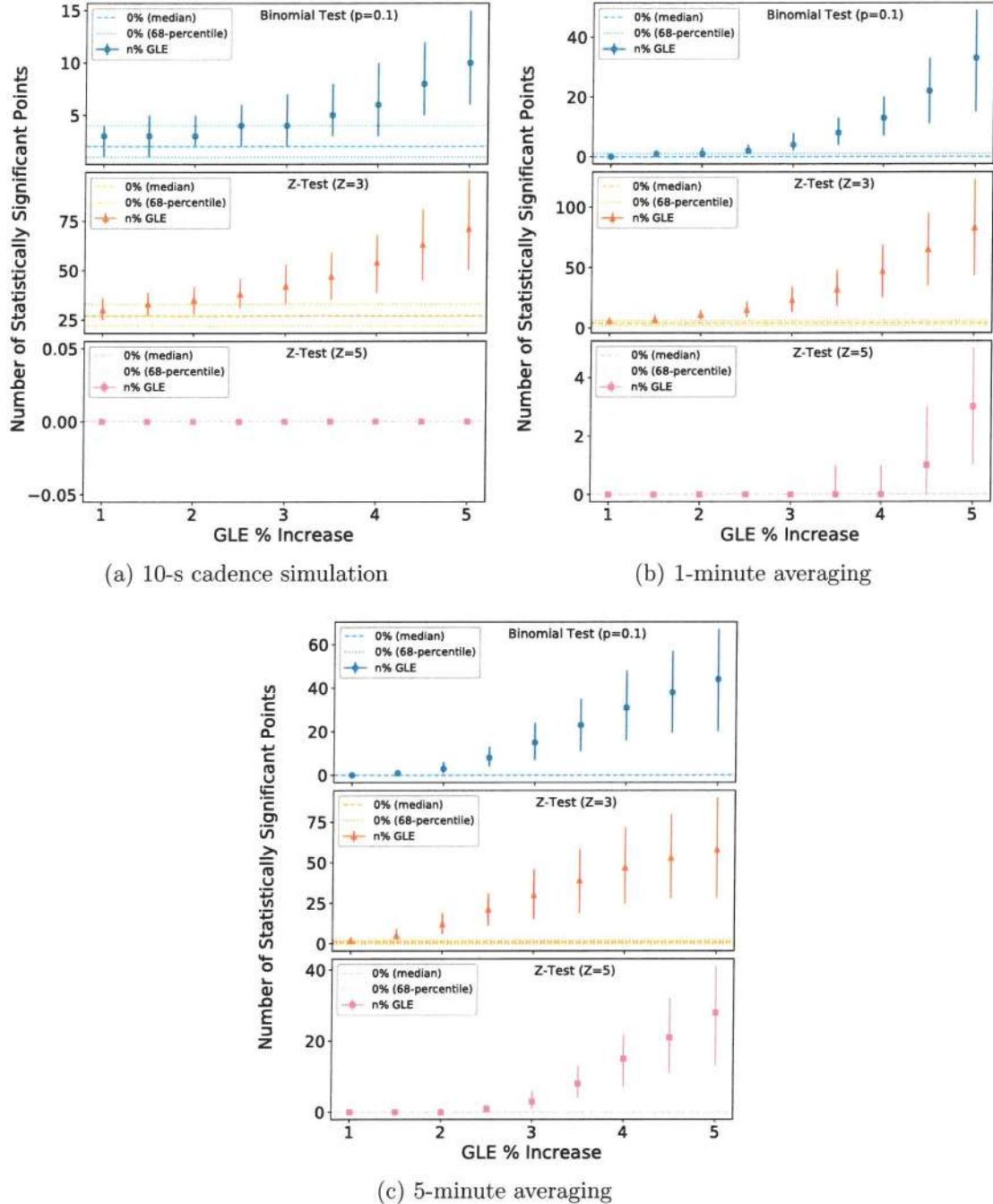


Figure 3.20: Summary plots showing the number of excessive measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for the 10-s cadence data; (b) 1-minute averaged data; (c) 5-minute averaged data. In each plot, the top panel shows the number of significant points exceeding the binomial $p = 10\%$ threshold, the middle panel shows the number of points exceeding the $Z=3$ threshold, and the bottom panel shows the number of points exceeding the $Z=5$ threshold. In each panel the dashed, horizontal lines show the median values of the tests without an injected GLE, and the horizontal, dotted lines represent the 68% credible intervals either side of the median. For each simulated GLE magnitude the point represents the median values and the error bars represent the 68% credible intervals either side of the median.

I would like to see the 7.5% and 10% case results somewhere. Either in Fig 3.20 (or similar) or at least in a table.

the $Z = 5$ significance level, there are no excessive measurements. In our complete analysis, we investigated GLE magnitudes up to 10 %, and determined that at the $Z = 5$ significance level we can differentiate the increase in the number of excessive events for GLE magnitude of $\geq 7.5 \%$. These limiting magnitudes are larger than typical magnitudes of GLEs predicted in Section 2.8.4, hence showing that despite reducing non-CR variations in the data, we are only capable of detecting the largest GLEs in the raw data acquired in this configuration.

When averaging the data over 1-minute and 5-minute intervals, our ability to detect lower-magnitude GLEs improved. We can differentiate GLEs from statistical fluctuations for magnitudes of 2.0–2.5 % and $\sim 1.5 \%$, for 1-minute and 5-minute averaging, respectively. Similarly, at the $Z = 5$ significance level this improved to 4.5–5.0 % and 2.5–3.0 %, respectively. Finally, using the binomial test, we expect to be able to differentiate the increase in the number of excessive measurements for a GLE with magnitude of $\sim 2.5 \%$ and $\sim 1.5 \%$, respectively.

We also note that with an increasing GLE magnitude we observe a larger spread in the average number of excessive points. This arises due to the random sampling of GLE properties, i.e. randomly sampling the rise and decay times. A GLE with a longer decay time generally results in a higher number of excessive points, compared to a GLE with a shorter decay time. This effect can be amplified when the magnitude of the GLE increases as the signal is more statistically significant. Therefore, we see an increase in the confidence intervals for larger GLE magnitudes in Figure 3.20.

This shows that through averaging the data, we expect, with a single detector, that we should be able to detect a GLE which induces an increase in the muon count rate on the order of 1.5–2 %. Weighing the benefits of the increased sensitivity against the timescales observable, we recommend making use of the 1-minute averaged data rather than 5-minutes, as the durations of some space weather events can be short-lived, which makes it advantageous to not to average over an ephemeral signal; however, there is added benefit in also analysing the 5-minute averaged data,

so a combination of all would be useful. In particular, the interactive GLE database tends to show data averaged over 5-minutes (Usoskin et al., 2016).

Overall, based on the values predicted in Section 2.8.4, we believe that it would have been possible to observe the increase in the muon count rate due to GLE 42, 43, and 45 with this new configuration. We have shown with the raw data we should be able to differentiate a GLE with magnitude $> 3.0\text{--}3.5\%$ (i.e. GLE 42) and when averaging the data into 5-minute bins, we expect to be able to observe GLEs with magnitudes $> 1.5\%$ (i.e. GLE 43 and 45).

3.6.4 Multiple Station Space Weather Uses

Many ground-based CR detectors typically exist as part of a network, which work together to increase their combined sensitivity. With an increasing number of stations in a network, observing the same events, the combined sensitivity increases by a factor of \sqrt{N} , where N is the number of stations in the network, due to the reduction in the Gaussian noise (although, of course, this is limited by the physical limitations of the detectors). However, it is also possible to use other methods to increase the sensitivity of the network and claim observations.

In this section we again use simulations of artificial data (see Appendix B for details) to determine the magnitude of GLEs that may be observed with a network of detectors using this configuration. One overarching assumption in this multi-station analysis was that the detectors are all geographically close, such that all the stations are triggered by the GLE at the same time, or the delay between the signals measured by individual detectors is negligible. We know this assumption is acceptable, based on the coincidental triggering of stations located across the world in the GNMN (Mishev & Usoskin, 2020). In addition, we assumed that all the stations had similar observing properties, as defined earlier. These assumptions are further supported if, for instance, we assume that we upgrade the HiSPARC stations closest to the University of Birmingham. There are another five operational stations in

Birmingham which are located within a radius of < 6 km from this new configuration, and a sixth station within a radius of ~ 15 km from University of Birmingham. With the exception of the existing University of Birmingham station 14001, which is a 4-detector station, each of the other stations in the Birmingham node have 2-detectors. This means that there are in total an additional 14 individual detectors in the Birmingham node which could be modified into this new configuration.

For each GLE magnitude we ran 1000 iterations of the simulations and were able to average over the number of excessive measurements for a given GLE magnitude. The GLEs were simulated with amplitudes of: 1.0%–5.0%; the start times, rise and decay times were all randomly sampled. This was done for 2-, 5-, and 10-stations. In each case we were able to perform the statistics tests on the mean of the data from all stations simulated. Table 3.5 shows the average number of measurements exceeding the various thresholds for simulations without an injected GLE.

Table 3.5: Median number of excessive measurements in the mean of the artificial data for multiple stations, with no injected GLEs. The values were acquired from 1000 iterations of simulated data, performing each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, with 1, 2, 5, and 10 stations. All uncertainties correspond to the 68% credible intervals either side of the median.

No. stations	No. excessive measurements		
	Binomial	$Z = 3$	$Z = 5$
1	2^{+2}_{-1}	27^{+6}_{-5}	0 ± 0
2	3^{+3}_{-2}	26^{+26}_{-13}	0 ± 0
5	2^{+4}_{-2}	24^{+30}_{-14}	0 ± 0
10	2^{+3}_{-2}	24^{+32}_{-15}	0 ± 0

To summarise the results of all the simulations, Figure 3.21 shows the average number of excessive measurements against the simulated GLE magnitude for the 10-s cadence observations, for 1, 2, 5, and 10 stations.

We see from Figure 3.21 that our ability to differentiate the number of excessive measurements for a GLE improves with a larger number of stations used. For a single station, using the binomial 10% threshold, we expect to be able to differentiate

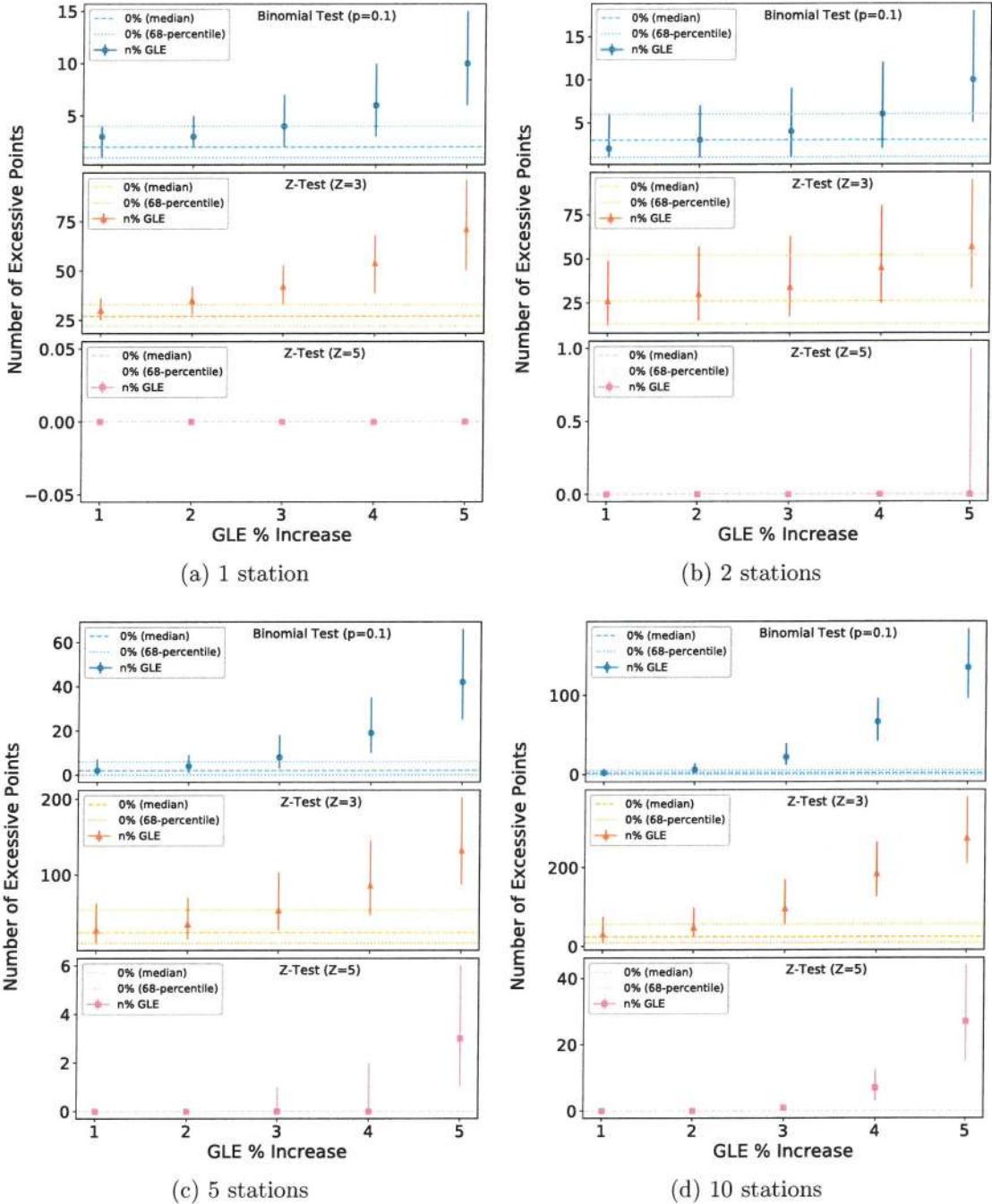


Figure 3.21: Summary plots showing the number of excessive measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for a single station; (b) 2 stations; (c) 5 stations; (d) 10 stations. In each plot, the top panel shows the number of excessive points exceeding the binomial $p = 10\%$ threshold, the middle panel shows the number of points exceeding the $Z=3$ threshold, and the bottom panel shows the number of points exceeding the $Z=5$ threshold. In each panel the dashed, horizontal lines show the median values of the tests without an injected GLE, and the horizontal, dotted lines represent the 68% credible intervals either side of the median. For each simulated GLE magnitude the point represents the median values and the error bars represent the 68% credible intervals either side of the median.

the increase in the number of excessive measurements for a GLE with magnitude of $> 4.0\text{--}5.0\%$. Increasing the number of stations to 5 and 10 improved the observable GLE magnitude to $\sim 3.0\text{--}4.0\%$ and $\sim 2.0\text{--}3.0\%$, respectively. In the case of combining 2 stations, we do not see an improvement due to an exasperation of statistical fluctuations. This shows a greater number of stations is necessary.

can
expand you
slightly? on this
why?

For a single station above the $Z = 3$ significance level, we have shown we are able to differentiate the increase in the number of excessive measurements for a GLE with magnitude of $\sim 3.0\text{--}4.0\%$; however, increasing the number of stations to 2, 5, and 10 did not improve our ability to observe lower magnitude events. Finally, we see using the $Z = 5$ significance level, there were no significant observations for magnitudes $\leq 5\%$ when only using a single station. In the multiple station analysis, we see that for 2, 5, and 10 stations this improved and the observable GLE magnitude reduced to $\sim 5.0\%$, $\sim 4.0\text{--}5.0\%$, and $\sim 3.0\text{--}4.0\%$, respectively.

Despite improving the sensitivity with an increasing number of stations, these limiting magnitudes are larger than typical increase in muon count rate due to GLEs that were shown in Section 2.8.4. This again shows that despite reducing non-CR variations in the data, we are only capable of detecting the largest GLEs in the raw data acquired in this configuration when directly interrogating the data.

As in Section 3.6.3, we again observe that increasing GLE magnitude increases the spread in the average number of excessive points. This again arises due to the random sampling of GLE properties, i.e. the rise and decay times, in the simulations leading to an increase in the confidence intervals for larger GLE magnitudes in Figure 3.21.

In addition to this analysis, we performed cross-correlation analyses between the stations simulated. This also relied on the assumption that the signal registered at each station has minimal delay, such that the peak of the Cross-Correlation Function (CCF) is at a time shift of zero. This analysis was performed for simulations of 2-, 5-, and 10-stations, with varying lengths of the GLE rise and decay. The resultant CCF

plots are shown in Figure 3.22, for a 2% magnitude GLE. The individual realisations of the CCFs, from combinations of pairs of stations, are plotted as black lines, while the red line shows the mean of the realisations.

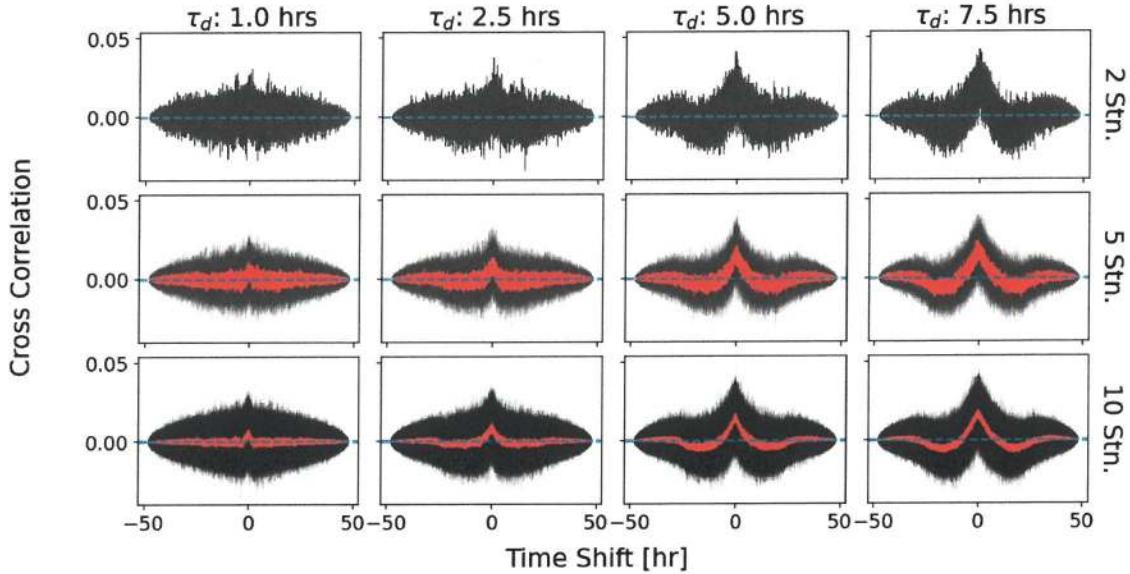


Figure 3.22: Cross-correlation analyses between 2, 5, and 10 stations for a 2% GLE magnitude with varying rise and decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the red line shows the mean of all the realisations. Finally, the dashed, horizontal lines depict a correlation of zero.

As the simulated data all experienced the GLE at the same time, there was a clear peak in the CCF at a time shift of zero hours, as expected, showing a strong asymmetry in the CCF around a zero-hour time shift. Assuming that a local network of stations using this configuration also experiences minimal delay between stations, we would expect to observe a similar CCF plot, allowing us to claim the detection of a GLE in each station. Figure 3.22 shows a strong dependence of the length of the GLE decay on the shape of the CCF, with a longer decay providing a clearer, broader CCF signal. The average decay time of GLEs as measured by Strauss et al. (2017) is $1.8^{+1.9}_{-1.3}$ hours, therefore few GLEs have decay times ≥ 5 hours. We should therefore expect that a ‘typical’ GLE with a similar magnitude would induce a CCF with a shape like the first or second columns, i.e. 1.0–2.5 hours.

We see from Figure 3.22 that increasing the number of stations means we can

average over the CCFs which results in a less-noisy CCF, shown by the red line. In the individual realisations of the CCFs (black lines) there is not a significant benefit at this level of GLE in increasing from 2 to 5 or 10 stations. However, the benefit of an increased number of stations is that we are able to reduce the noise on the combined CCF signal. This is advantageous as it allows us to more clearly detect the correlated signals between 5 and 10 stations, versus with only 2 stations. For the simulation using two stations and a decay time of 1 hour in Figure 3.22, it is difficult to determine a peak near zero-hour time shift, but increasing to 5 and 10 stations shows the benefit, as in the mean CCF we then see the peak at a zero-hour time shift.

This analysis was repeated for simulations of a 1 % GLE, to investigate whether the increased number of stations allow us to observe GLEs with such a low magnitude. The results are shown in Figure 3.23.

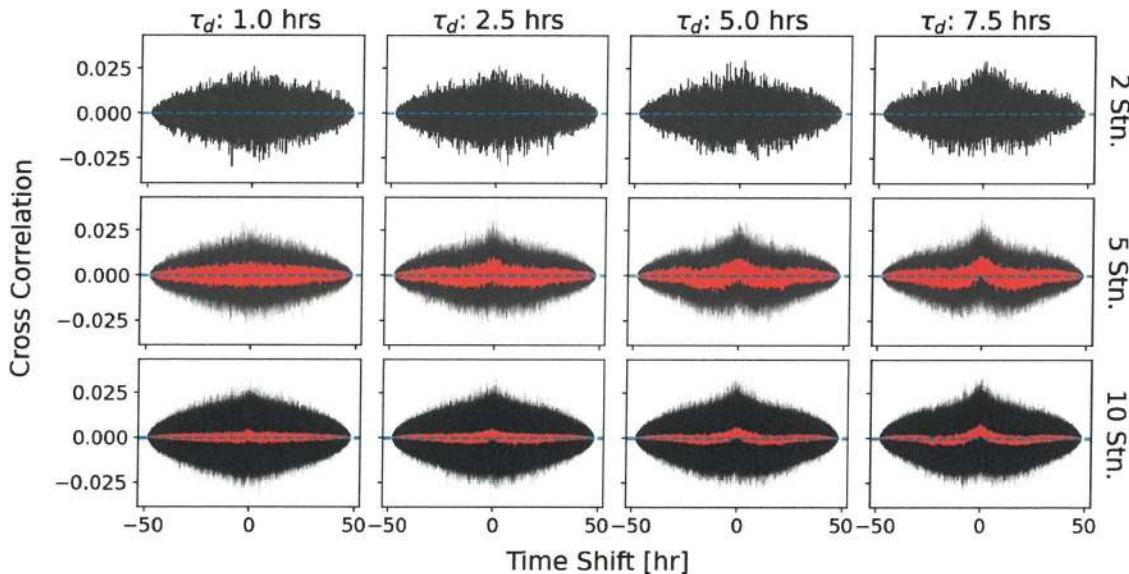


Figure 3.23: Cross-correlation analyses between 2, 5, and 10 stations for a 1% GLE magnitude with varying rise and decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the red line shows the mean of all the realisations. Finally, the dashed, horizontal lines depict a correlation of zero.

With a 1 % GLE magnitude it becomes even harder to determine the zero-hour peak. For the simulations with a decay time of 1 hour in Figure 3.23, it is difficult to

determine any peak near zero-hour time shift for the case with 2 and 5 stations and it becomes only slightly visible when increasing to 10 stations. On the other hand, for a decay time of 7.5 hours we can clearly see the CCF shape with a peak at zero hours in all cases, i.e. using 2-, 5-, and 10- stations, but we know the average GLE decay time is closer to 2 hours (Strauss et al., 2017). At the 2.5 hour decay scale we see the peak in both the 5- and 10-station CCFs is observable. Figure 3.23 therefore shows us that at the 1 % GLE scale, we should sensibly increase the network to 5- or 10-stations to ensure the best chance of observing the cross-correlation in the mean CCF.

We generally expect the magnitude of the muon increase due to GLEs is on the order of or less than $\sim 1 \%$, based on the results shown in Section 2.8.4. We have shown that we do not expect to observe GLEs with a single HiSPARC Muon Detector (MD). We therefore recommend here that any upgrades to form a network should ideally use 10 stations and no fewer than 5 stations, in order to be able to resolve the cross-correlation between the stations.

3.7 Conclusions

In this chapter we have presented a new HiSPARC station configuration and investigated its performance for use monitoring space weather events. We have outlined the set-up of the new station, from the configuration of the scintillators, calibration of the PMTs, processing of the PMT signals using a series of NIM modules, and the acquisition of data using a Raspberry Pi. We have also outlined how we acquire atmospheric pressure data and the temperature measured within the roof box. Both measurements were used to correct for effects on the acquired CR data.

The station was configured to collect two types of coincidence data: (i) the full coincidence counts between the two PMTs; (ii) a reduced coincidence count, which is between the original coincidences and a gate signal. The reduced count rate was necessary to not overload the HiSPARC servers.

The removal of the atmospheric effects was routinely performed on the acquired singles and coincidences data. We showed for the coincidences data it was only necessary to correct for the effects of pressure. The configuration of the station uses the coincident signal from two separate PMTs and we showed there was only a weak correlation with temperature which was likely due to the diurnal CR anisotropy rather than temperature. This was further supported in our investigation of the spurious counts, when investigating the noise in the coincidences data, which did not show a diurnal signal. The temperature correction was however still necessary for singles data and we demonstrated the success of this method. In both data it was necessary to correct for atmospheric pressure, which we showed was successful when using pressure data measured \sim 20 km from this HiSPARC station.

After atmospheric corrections, we analysed the coincidences data by using a Poisson likelihood function to determine the posterior distribution for the mean of the count rate. In addition, it was demonstrated by adding a large delay between the two PMTs that we were able to quantify the noise (spurious counts) in the coincidences data, caused by random coincidences between the two PMTs. As discussed above, this noise did not show any diurnal variation, meaning the diurnal temperature effects in the singles data did not bleed through into the coincidences data. Through the analysis of the original and reduced coincidences data, we were able to quantify the reduction factor, which was as expected from the duty cycle of the gate signal and we showed that there is agreement between the reduced data stored locally and that stored on the HiSPARC servers, with only differences being realisations of the Poisson noise.

A further study was conducted to understand the magnitudes of GLEs that should be observable with this new configuration. This used simulation of artificial data and we were able to perform statistical tests, comparing the number of statistically significant measurements in 2 days of data with GLEs of varying magnitude. We showed this method was suitable for claiming observations of GLEs and showed

that through averaging the data over 1- and 5-minutes, we can reduce the noise to observe lower GLE magnitudes.

Expanding on this, we also performed analyses using artificial data with multiple stations. This was done to examine whether upgrading to a network of several HiSPARC stations in this configuration would improve our ability to observe low-magnitude GLEs. We performed statistical tests on the mean signal between the stations and also performed cross correlation analyses. We demonstrated that there exists a strong dependence on the decay time of the GLE on the shape of the CCF signal, but increasing the number of stations allowed us to observe the correlation.

We leave the reader with the following points:

1. A new configuration of HiSPARC detector has been implemented, which is more relevant for space weather applications, as it overcomes the bias towards higher energy PCRs that the HiSPARC experiment intends to observe. It also reduces the atmosphere induced diurnal effects.
2. The mean count rate in this configuration is $\sim 80 \mu/\text{s}$ and the noise from spurious counts is of about $0.0043 \pm 0.0002 \mu/\text{s}$. This noise is small and negligible compared to the Poisson noise of $\sim 9 \mu/\text{s}$, which represents $\sim 11\%$ of the signal. The reduced coincidences data has a count rate which is lower by a factor of approximately 110, and the reduced counts data has been shown to be in agreement for the data stored locally and that stored in the HiSPARC servers.
3. There exists a good visual agreement between the data monitored by this station and a local NM station, Dourbes, in Belgium. The data from this station and from the Dourbes NM station should be continually compared to monitor this relationship, as it will be instrumental when or if a space weather event occurs in the next Solar Cycle.
4. Simulations of artificial data demonstrated that with the raw, 10-s cadence

observations we should expect to be able to detect GLEs with this configuration with a magnitude of $\gtrsim 3\text{--}4\%$. Through averaging the data into 1- or 5-minute bins, we reduce the noise and improve the sensitivity to observe GLEs with magnitudes $\gtrsim 1.5\text{--}2.0\%$. This is in line with some of the predicted GLE magnitudes from Chapter 2.

5. Through simulating the performance of a network of detectors in this configuration we showed that we can improve the sensitivity, to observe GLEs on the order of $\sim 1\%$, through analysing the cross-correlation of nearby stations. However, we note that there is a strong dependence on decay time of the GLEs. We recommend that any upgrades to form a network should ideally use 10 stations and no fewer than 5 stations.
6. It is important that this configuration of HiSPARC detector is maintained until at least 2026, to ensure a complete study is performed to at least the maximum of Solar Cycle 25.
(expected)

If there a plan
in place for this?

4 Galactic Cosmic Ray Behaviour During Solar Cycle 24

The majority of the text in this chapter is taken verbatim from Ross & Chaplin (2019). I was first author of the journal article and conducted the entirety of the work in the investigation. Section 4.6 has been written since the publication to provide a comparative analysis using the High School Project on Astrophysics and Research with Cosmics (HiSPARC) data.

4.1 Introduction

Galactic Cosmic Rays (GCRs) are charged particles and atomic nuclei with energies spanning the range from a few MeV up to approximately 10^{21} eV, that encroach upon the Earth from all directions (Giacalone, 2010). They mainly originate outside the solar system, within the Milky Way; however, they are also expected to originate from other galaxies (Aab et al., 2017). GCRs at the top of the atmosphere are mostly composed of protons ($\sim 87\%$) and α -particles ($\sim 12\%$), with a smaller contribution ($\sim 1\%$) from heavier nuclei (Dunai, 2010).

When Cosmic Rays (CRs) enter the atmosphere, they interact with atmospheric atoms and produce cascades of secondary particles, which at ground level are primarily neutrons and muons. Neutron Monitors (NMs) and Muon Detectors (MDs) located at different locations on Earth have been used since the 1950s to observe GCRs. Information on GCRs prior to the modern epoch of NMs and MDs, and the

4 Galactic Cosmic Ray Behaviour During Solar Cycle 24

The majority of the text in this chapter is taken verbatim from Ross & Chaplin (2019). I was first author of the journal article and conducted the entirety of the work in the investigation. Section 4.6 has been written since the publication to provide a comparative analysis using the High School Project on Astrophysics and Research with Cosmics (HiSPARC) data.

4.1 Introduction

Galactic Cosmic Rays (GCRs) are charged particles and atomic nuclei with energies spanning the range from a few MeV up to approximately 10^{21} eV, that encroach upon the Earth from all directions (Giacalone, 2010). They mainly originate outside the solar system, within the Milky Way; however, they are also expected to originate from other galaxies (Aab et al., 2017). GCRs at the top of the atmosphere are mostly composed of protons ($\sim 87\%$) and α -particles ($\sim 12\%$), with a smaller contribution ($\sim 1\%$) from heavier nuclei (Dunai, 2010).

When Cosmic Rays (CRs) enter the atmosphere, they interact with atmospheric atoms and produce cascades of secondary particles, which at ground level are primarily neutrons and muons. Neutron Monitors (NMs) and Muon Detectors (MDs) located at different locations on Earth have been used since the 1950s to observe GCRs. Information on GCRs prior to the modern epoch of NMs and MDs, and the

observations we should expect to be able to detect GLEs with this configuration with a magnitude of $\gtrsim 3\text{--}4\%$. Through averaging the data into 1- or 5-minute bins, we reduce the noise and improve the sensitivity to observe GLEs with magnitudes $\gtrsim 1.5\text{--}2.0\%$. This is in line with some of the predicted GLE magnitudes from Chapter 2.

5. Through simulating the performance of a network of detectors in this configuration we showed that we can improve the sensitivity, to observe GLEs on the order of $\sim 1\%$, through analysing the cross-correlation of nearby stations. However, we note that there is a strong dependence on decay time of the GLEs. We recommend that any upgrades to form a network should ideally use 10 stations and no fewer than 5 stations.
6. It is important that this configuration of HiSPARC detector is maintained until at least 2026, to ensure a complete study is performed to at least the maximum of Solar Cycle 25. (expected)

*Is there a plan
in place for this?*

space age, rely on the studies of cosmogenic isotope records from ice cores and tree rings (Owens & Forsyth, 2013).

It has long been established that there exists an anti-correlation between GCR intensity and the level of solar activity, over a cyclic 11-year period, with perhaps some time-lag (Forbush, 1958; Parker, 1965; Usoskin et al., 1998; Van Allen, 2000). Figure 4.1 shows clearly the anti-correlation between GCRs and Sun Spot Number (SSN).

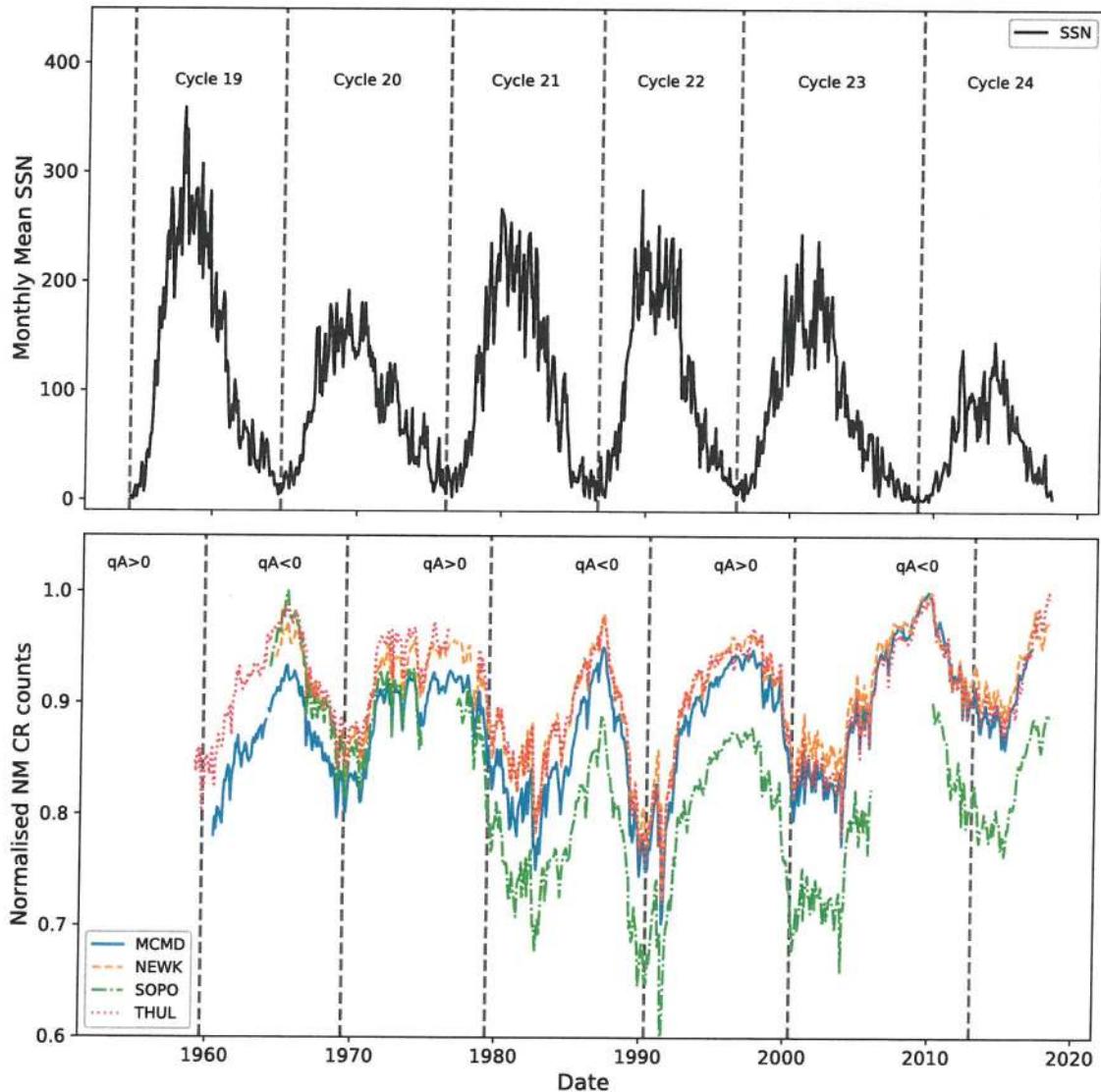


Figure 4.1: SSN (top), with vertical lines showing the beginning of each solar cycle. Cosmic ray intensity recorded by NMs (bottom), with vertical lines showing the approximate epochs of solar magnetic field polarity reversals. (MCMD = McMurdo, NEWK = Newark, SOPO = South Pole, THUL = Thule).

It is well known that the 11-year solar activity cycle is in fact a 22-year cycle – the Hale cycle – which describes the alternating polarity of the large-scale solar magnetic field (Thomas et al., 2014a; Hathaway, 2015). The interchanging peaked and flat-topped shape of GCR intensity in Figure 4.1 is a manifestation of this effect in addition to other CR transport processes (Aslam & Badruddin, 2012).

The polarity of the solar field, A , is taken to be positive when the polar field is outward in the northern hemisphere, i.e. aligned with the axis of rotation, and negative when the opposite is true (Thomas et al., 2014a). The solar field polarity conventionally is described in combination with particle charge, q , due to the effect of curvature and gradient drift on charged particles; thus it is customary to define the solar polarity as qA . Vertical lines showing the approximate epochs at which the polarity reverses are plotted in Figure 4.1 (Janardhan et al., 2018; Thomas et al., 2014a).

Particle drifts differ during different qA cycles, with positive CRs (i.e. protons) predominantly arriving into the heliosphere from the heliospheric poles and outwards to Earth during periods when $qA > 0$, whereas when $qA < 0$, positive CRs predominantly arrive at Earth inwards along the Heliospheric Current Sheet (HCS) (Belov, 2000; Thomas et al., 2014b). As the solar magnetic dipole axis is tilted to the solar rotation axis, so is the HCS; the HCS tilt varies with solar cycle and is typically smaller during solar minimum and larger during solar maximum (Owens & Forsyth, 2013). The tilt angle of the HCS has also been shown to be strongly correlated to the GCR intensity and the GCR lag behind the solar activity (Belov, 2000; Mavromichalaki et al., 2007).

Aslam & Badruddin (2012, 2015) found that the different processes of CR transport have varying levels of importance throughout the solar activity cycle, but around solar maximum it is likely that drifts play less of a role and disturbances in the solar wind (and hence HCS) are the predominant factor of CR modulation. Even cycles encounter $qA < 0$ polarity during their onset phase and $qA > 0$ during

their declining phase, thus experiencing a faster GCR recovery after solar maximum as the GCRs predominantly enter the heliosphere from the heliospheric poles and experience an outwards drift towards Earth. Odd cycles encounter $qA > 0$ polarity during their onset phase and $qA < 0$ during their declining phase and so experience a slower recovery after solar maximum, as the GCRs predominantly enter the heliosphere along the HCS. When the HCS is tilted and disturbed during the declining activity phase, the path length that GCRs must travel to Earth increases; hence resulting in an increased time-lag.

Several studies have demonstrated the lag between GCR and solar activity proxies is approximately zero (i.e. no lag) during even solar cycles, and that there exists a lag of around a year or more during odd solar cycles (Usoskin et al., 1998; Mavromichalaki et al., 2007; Singh et al., 2008).

[Van Allen \(2000\)](#) showed through cross-plotting the annual mean intensity of GCRs against SSN between 1953 and 1999 (covering solar cycles 19–22), that there is a distinct difference in the plot shapes between the different solar cycles, with 19 and 21 producing broad ovals, and 20 and 22 as approximately flat lines. The striking difference between odd-numbered and even-numbered cycles is shown for cycles 19–24 in Figure 4.2. It is believed that this hysteresis effect is caused by the combination of the Heliospheric Magnetic Field (HMF), solar magnetic field polarity and thus the particle drift, and the tilt of the HCS leading to a slow recovery of GCR intensity after maxima in odd cycles and a fast recovery after maxima in even cycles ([Van Allen, 2000](#); [Belov, 2000](#); [Thomas et al., 2014a](#)).

An extension of this work has since been carried out by [Inceoglu et al. \(2014\)](#) showing that the even numbered solar activity cycles can be best modelled using a linear fit due to the narrow shape of the hysteresis loops; whereas odd-numbered solar activity cycles are better represented by ellipses due to their broader shape.

There has been speculation in the literature on the behaviour of cycle 24 compared to recent odd and even cycles. It has been suggested that there exists a lag

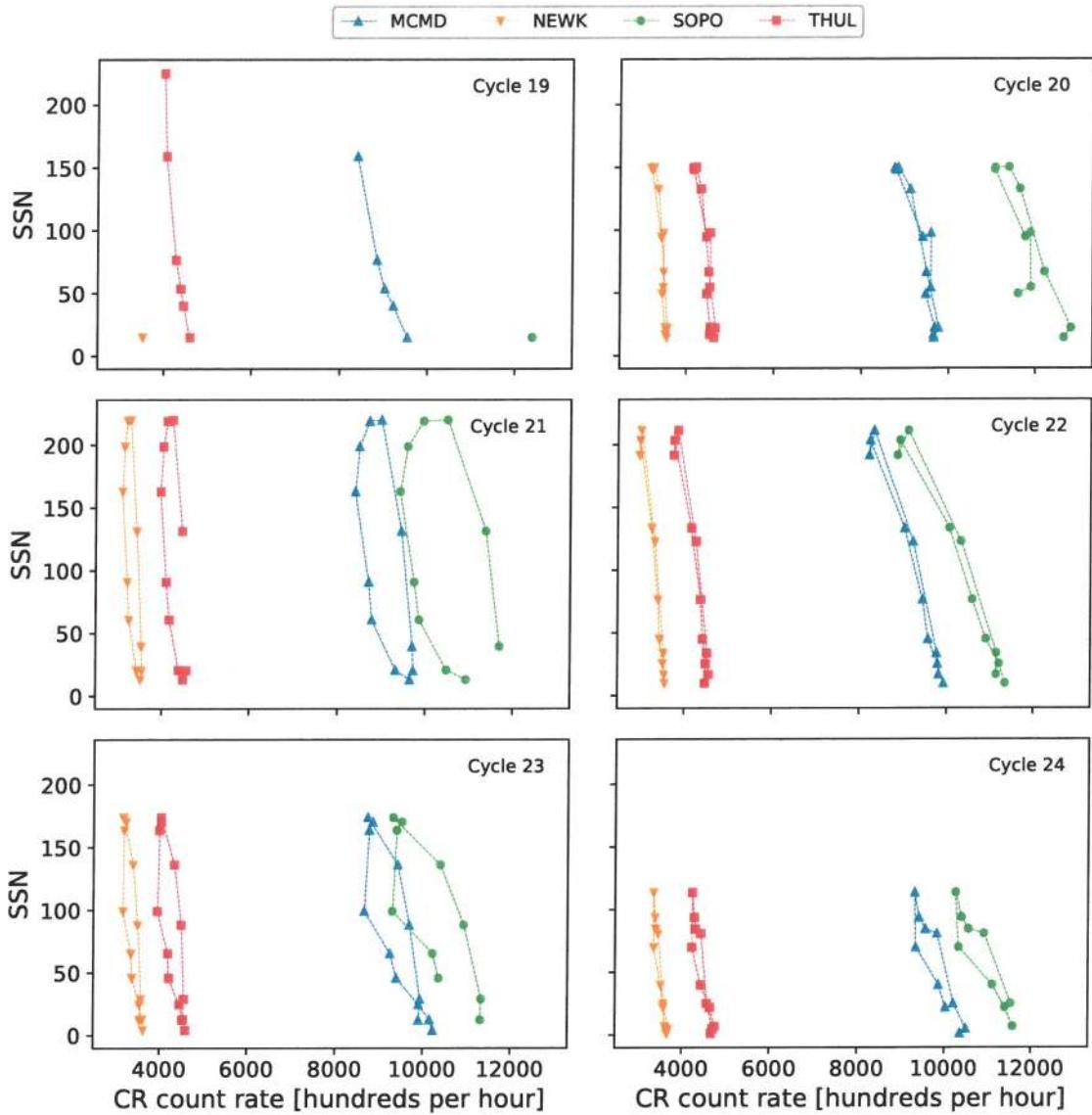


Figure 4.2: Hysteresis plots between yearly averaged SSN and yearly averaged GCR intensity for each of the 4 main NM stations over cycles 19–24.

between SSN maxima and GCR intensity minima in excess of 10 months (Kane, 2014; Mishra & Mishra, 2016) which does not follow the previous even cycles having a near-zero lag and in fact suggesting that cycle 24 behaved similarly to previous odd cycles; however, these studies do not make use of a complete cycle of data and thus may draw inaccurate conclusions about the behaviour of the whole cycle because of the unusually extended nature of the declining phase of cycle 23 and the low amplitude of cycle 24 maximum (Broomhall, 2017). Mishra & Mishra (2016) make use of a more complete data set for cycle 24, yet still incomplete, and conclude

that it is also likely that a 4 month lag could exist between GCR and SSN.

This work aims to provide a timely update on the statistical relationship between GCRs intensity and solar activity during solar cycle 24, since the cycle has now almost declined to a minimum. These aims have been achieved through a time-lag analysis and hysteresis effect analysis between SSN and GCR intensity.

In Section 4.2 we provide a brief description of the data that was used throughout this study for both CRs and SSN.

We show in Section 4.3 through a correlative time-lag analysis that there exists a small time-lag between the SSN and GCR intensity over solar cycle 24, which is slightly longer than preceding even-numbered cycles but not as high as observed in previous odd-numbered cycles. We also discuss whether the time-lag between SSN and GCR shows a dependence on the rigidity cut-off of the observing station.

In Section 4.4 we model the shapes of hysteresis plots between GCR intensity and SSN. We show that the behaviour of the hysteresis loops for cycle 24 follow the preceding even-numbered solar activity cycles and is better represented by a straight line fit rather than an elliptical model.

4.2 Data

For the majority of the work in this study we have considered the pressure corrected count rates measured by four NM monitor stations as acquired from the Neutron Monitor Data Base (NMDB) ([NMDB, 2018](#)) event search tool (NEST) (<http://nmdb.eu/nest/>). The four stations are McMurdo (MCMD), Newark (NEWK), South Pole (SOPO), and Thule (THUL), i.e. the same NM stations used in the study by [Inceoglu et al. \(2014\)](#) to provide a comparison to existing literature. Table 4.1 details the basic characteristics of the NM stations used in this study.

We have investigated the long-term GCR modulation in the heliosphere from 1964–2018, spanning solar cycles 20–24, for the cycle epochs: 20: (10/1964 – 03/1976); 21: (03/1976 – 09/1986); 22: (09/1986 – 05/1996); 23: (05/1996 –

12/2008); 24: (12/2008 – 03/2018). Early predictions on solar cycle 25 suggest that solar cycle 24 is unlikely to reach a minimum earlier than the middle of 2019 up to as far as early 2021 (see Howe et al. (2018); Upton & Hathaway (2018); Pesnell & Schatten (2018)). The data used in this study are therefore of an incomplete cycle 24; however, we believe this to now have a minimal effect on the results as cycle 24 draws to a minimum. Cycle 19 was omitted from this study due to the incomplete data set for this period (see Figure 4.1 and Figure 4.2).

Table 4.1: Neutron monitor stations used in this study and their vertical geomagnetic cut-off rigidity (R_c), longitude, latitude, and altitude acquired from NEST. The first four stations have been used for all of the analysis while the lower 12 stations have been used exclusively for the investigation into the dependence of R_c on the time-lag.

	Station	R_c [GV]	Long. [deg]	Lat. [deg]	h [m]
Time-Lag & Hysteresis	McMurdo (MCMD)	0.30	166.6 E	77.9 S	48
	Newark (NEWK)	2.40	75.8 W	39.7 N	50
	South Pole (SOPO)	0.10	0.0 E	90.0 S	2820
	Thule (THUL)	0.30	68.7 W	76.5 N	26
	Oulu (OULU)	0.81	25.5 E	65.1 N	15
	Kerguelen (KERG)	1.14	70.3 E	49.4 S	33
	Magadan (MGDN)	2.10	151.1 E	60.0 N	220
	Climax (CLMX)	3.00	106.2 W	39.4 N	3400
	Dourbes (DRBS)	3.18	4.6 E	50.1 N	225
	IGY Jungfraujoch (JUNG)	4.49	7.98 E	46.6 N	3570
	Hermanus (HRMS)	4.58	19.2 E	34.4 S	26
	Alma-Ata B (AATB)	6.69	76.9 E	43.0 N	3340
R_c dependence of time-lag	Potchefstroom (PTFM)	6.98	27.1 E	26.7 S	1351
	Mexico (MXCO)	8.28	99.2 W	19.8 N	2274
	Tsumeb (TSMB)	9.15	17.6 E	19.2 S	1240
	Huancayo (HUAN)	12.92	75.3 W	12.0 S	3400

During the time-lag correlation analysis, as our results suggested there may be a rigidity dependence on the time lag, we introduced a further 12 NM stations with data acquired from NEST spanning cycles 20–24 to increase the rigidity spectrum utilised in this study; these stations and their basic characteristics are also detailed in Table 4.1. These stations are not included in the rest of the results however, as the results from these stations do not change the conclusions of this study.

Furthermore, we have also used monthly/yearly averaged SSN, as collected by WDC-SILSO (<http://sidc.be/silso/>), for the time-lag analysis/hysteresis analysis respectively as our chosen proxy of solar activity.

4.3 Time-Lag Analysis

4.3.1 Method

To investigate the time delay between the modulation of GCRs compared to the solar activity, a time-lag cross-correlation analysis was performed between monthly mean GCR intensity and monthly mean SSN for each station, following the approach of Usoskin et al. (1998). We used a time window of width T centred on a time t , i.e. shifting within the interval $t - T/2$ to $t + T/2$. Here we used $T = 50$ months.

The window was shifted in steps $\Delta t = 1$ month within this interval and for each step the Spearman's rank correlation coefficient (ρ) between GCR intensity and SSN was calculated. The lag between GCR and SSN was then estimated by finding the peak correlation coefficient within the time interval T .

The results from the four main NM stations used suggested that there may be a relationship between the rigidity cut-off (R_c) of a NM station and the time-lag for GCRs; hence four additional NM stations were introduced to determine whether this was so, as detailed above.

4.3.2 Results

The correlation (ρ) between monthly averaged GCR counts and SSN for different time-lags was calculated for cycles 20–23. The variation in ρ is presented in Figure 4.3, showing that for each cycle there is a time-lag corresponding to peak anti-correlation between GCR intensity and SSN. Table 4.2 summarises the time-lag with the highest correlation and the corresponding correlation coefficient for all stations in each individual solar cycle.

As previously reported in the literature, we see here that all of the NM stations

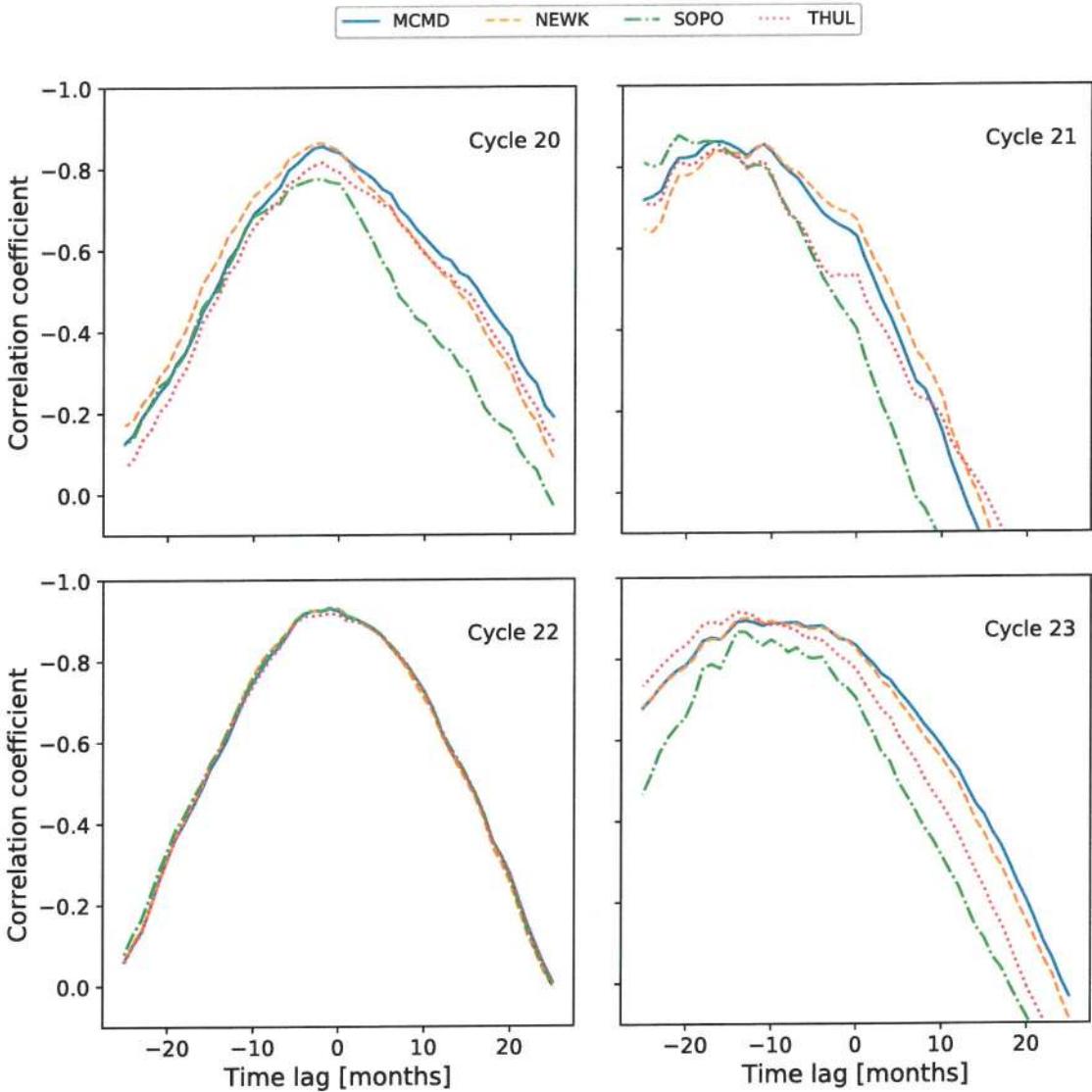


Figure 4.3: Variation in the correlation coefficient with time-lag NM station GCR intensity and SSN during solar cycles 20–23.

clearly exhibit almost no lag during even solar cycles, and a longer lag varying between 11–21 months during odd solar cycles. There is a strong agreement between the results presented in Table 4.2 and those of Mavromichalaki et al. (2007), Kane (2014), and Paouris et al. (2015), thus providing further evidence on the distinction between odd and even solar cycles due to particle transport in the heliosphere. The agreement with existing literature provides evidence of a suitable methodology in this study.

The same cross-correlation technique was then applied to cycle 24 between the

Table 4.2: Time-lags and the corresponding cross-correlation coefficient between NM CR count and SSN for solar cycles 20–23.

	Cycle 20		Cycle 21	
	Lag [months]	ρ	Lag [months]	ρ
McMurdo	2	-0.855	16	-0.862
Newark	2	-0.863	11	-0.856
South Pole	3	-0.776	21	-0.877
Thule	2	-0.816	17	-0.841
	Cycle 22		Cycle 23	
	Lag [months]	ρ	Lag [months]	ρ
McMurdo	1	-0.929	13	-0.894
Newark	1	-0.931	13	-0.900
South Pole	1	-0.931	14	-0.866
Thule	1	-0.917	14	-0.914

Perhaps something to do with the effect post 1980 in South Pole Fig 4.1

dates 12/2008 – 03/2018 and the results are presented in Figure 4.4 and Table 4.3.

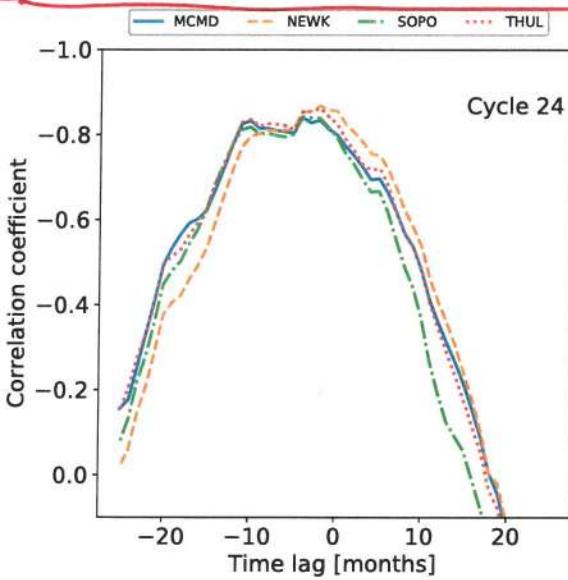


Figure 4.4: Variation in the correlation coefficient with time-lag between NM GCR intensity and SSN during solar cycle 24.

Table 4.3: Time-lags and the corresponding cross-correlation coefficient between NM GCR intensity and SSN for solar cycle 24.

	Cycle 24	
	Lag [months]	ρ
MCMD	4	-0.841
NEWK	2	-0.868
SOPO	4	-0.843
THUL	2	-0.862

Cycle 24 is seen here to follow the pattern of almost no lag for even cycles;

however, cycle 24 does display a lag that is larger than the previous two even-numbered cycles, despite not being as large as the two previous odd-numbered cycles.

The cause for the increased time-lag in cycle 24, as compared to the previous two even-numbered cycles, is likely due to the combined effects of the unusually deep

and extended minimum between solar cycles 23 and 24, which delayed the decline in GCR intensity and caused record-breaking high GCR intensities (Pacini & Usoskin, 2015), and the small amplitude of the cycle 24 maximum.

The results presented in this study, using data for a near-complete cycle 24, show that the results of Kane (2014), and Mishra & Mishra (2016), were likely unduly influenced by the unusually deep and extended declining phase of cycle 23 given that they had a limited data set. Mishra & Mishra (2016) used data for just over half of cycle 24 and resulted in a time-lag of 4 months which agree with the results of this study.

As a further note on time-lag, Tomassetti et al. (2017) showed that through the introduction of time-lag as a parameter in the CR transport calculations of CR spectra that there exists a time-lag of 8.1 ± 1.2 months during the period 2000–2012 spanning across cycle 23 and 24. We performed the time-lag analysis for the first 4 NM stations detailed in Table 4.1 for the period between 2000–2012 to investigate whether these results can be reproduced. The results of this analysis are presented in Figure 4.5 and Table 4.4.

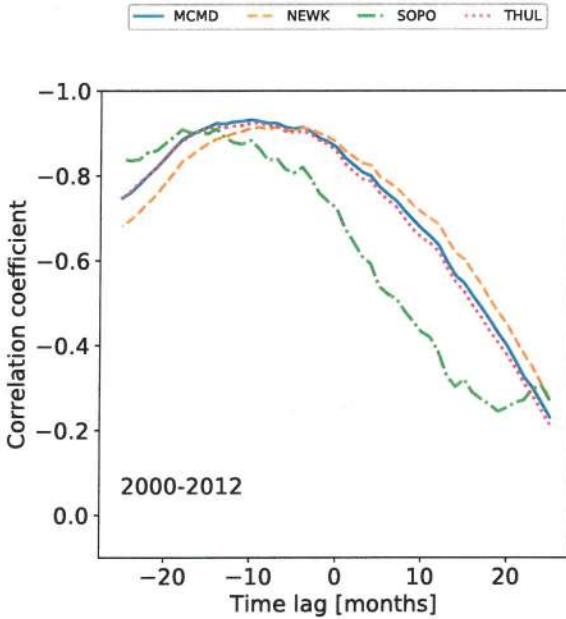


Table 4.4: Time-lags and the corresponding cross-correlation coefficient between NM GCR intensity and SSN during 2000–2012.

2000–2012		
	Lag [months]	ρ
MCMD	10	-0.932
NEWK	7	-0.914
SOPO	14	-0.911
THUL	9	-0.925

Figure 4.5: Variation in the correlation coefficient with time-lag between NM GCR intensity and SSN between 2000–2012.

From the time-lag analysis of the 4 stations presented there is a mean lag of 10.00 ± 1.47 months, which is in good agreement with the results of Tomassetti et al. (2017).

Finally, allowing for the odd/even cycle dependence, we see in Figure 4.3, Figure 4.4, and Figure 4.5 that the time-lag appears to depend on the rigidity of the NM station used for observation. Such a dependence may impact the conclusions depending on the choice of NM station. We expect that if a dependence exists, a station with a higher rigidity cut-off (R_c) would have a shorter lag as this station observes higher energy CRs which are affected less by solar modulation and thus able to recover faster from solar maximum. Whereas a station with a lower cut-off rigidity observing lower energy CRs, which are more influenced by solar modulation, would recover more slowly from solar modulation and therefore experience a longer time-lag. This is supported by Figure 4.3, Figure 4.4, and Figure 4.5, but in order to provide more conclusive evidence of such a relationship we introduced the additional NM stations detailed in Table 4.1 to provide a rigidity range spanning 0–13 GV. We present in Figure 4.6 a plot of the time-lag versus station R_c for all 16 stations over cycles 20–24. To acquire uncertainties on the time-lag we ran 1000 Monte Carlo simulations of the time-lag analysis, sampling from the uncertainty distributions for each of the monthly averaged SSN and GCR counts; however, the uncertainties in the data propagated in the Monte Carlo simulations produced no scatter in the overall results.

The results of this analysis do not show a clear rigidity dependence on the time-lag between SSN and GCR intensity; the sampling of higher R_c is too low, due to the availability of high R_c stations, to reasonably conclude on such a dependence at high rigidities, despite cycle 21 suggesting the existence of a dependence for low R_c stations as per our expectations. For low R_c stations there appears to be a more pronounced distinction between the time-lag observed between odd-numbered and even-numbered cycles than for higher R_c stations; however, again this is not

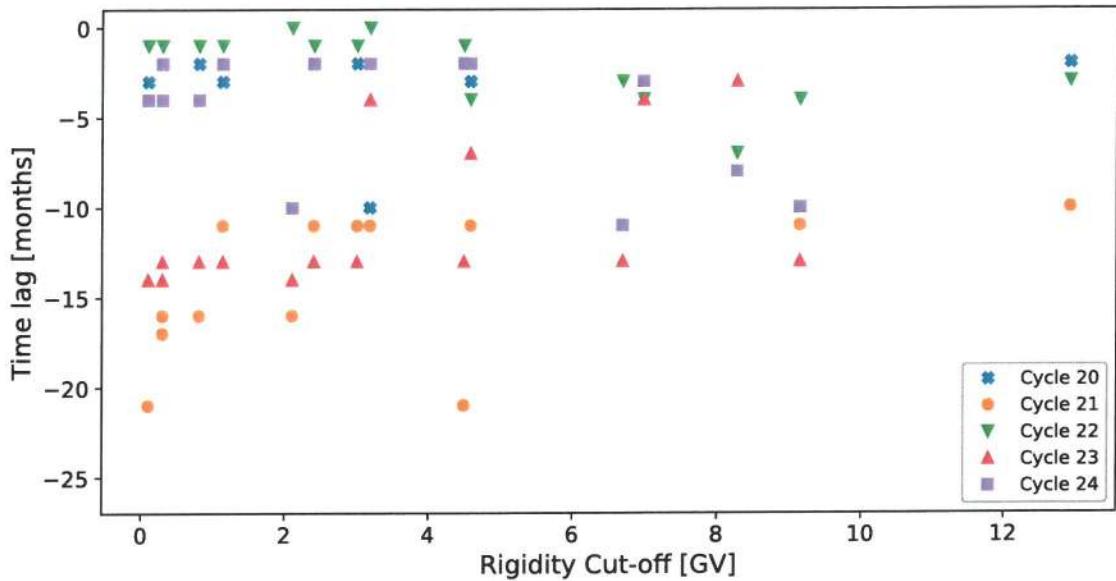


Figure 4.6: Variation in time-lag plotted against NM station rigidity cut-off for the 16 NM stations detailed in Table 4.1.

definitive due to the low sampling at higher R_c . We therefore conclude that there will be no significant dependence of the time-lag analysis on the R_c of the observing station.

4.4 Hysteresis Effect Analysis

4.4.1 Method

To investigate the hysteresis effect, we have adopted the approaches of Van Allen (2000), Singh et al. (2008), and Inceoglu et al. (2014). Plots of the annual mean SSN versus the annual mean GCR intensity were generated for cycle 20–24 and analysed by fitting different models to the data.

As highlighted in Inceoglu et al. (2014), even-numbered solar cycles can be suitably modelled by a linear fit due to their narrow hysteresis shape, and odd-numbered solar cycles were shown to be better modelled by ellipses due to their broadened hysteresis shape. Here, we first repeated, for solar cycles 20–23, the linear and ellipse fitting to confirm the method reproduces the results reported in Inceoglu et al. (2014) before applying the method to cycle 24.

For even cycles which display narrow hysteresis loops, an unweighted least squares linear regression was used to reconstruct estimates of the GCR intensity from SSN. As odd-numbered solar cycles display a broader hysteresis loop, they were separately modelled using unweighted linear regression and ellipse fitting to determine the model providing the better fit.

The equation of the ellipse fitting took the form:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + R(\phi) \begin{bmatrix} a \cos \theta \\ b \sin \theta \end{bmatrix}, \quad (4.1)$$

where x is GCR intensity; y is SSN; (x_0, y_0) are the centroid coordinates of the fitted ellipse, $R(\phi)$ is the rotation matrix; ϕ is the ellipse tilt angle; a and b are the semi-major and semi-minor axes respectively; and $0 \leq \theta \leq 2\pi$ is the polar angle measured anti-clockwise from the semi-major axis.

In order to regain the GCR intensity from the model, where linear regression was used to model the data, GCR was acquired directly from the SSN for each year. For the ellipse model the GCR was acquired from the model as a function of time from θ , where the time-lag calculated from the analysis in Section 4.3 was used to correctly phase the ellipse allowing θ to be calculated using standard equations of ellipses.

GCR intensity predicted by the linear regression and ellipse models were compared to the measured GCR intensity using Spearman's rank correlation as per Inceoglu et al. (2014).

4.4.2 Results

The hysteresis loops between yearly averaged SSN and GCR intensity for each station were first modelled with a linear regression for both odd and even solar activity cycles, then the odd cycles were separately re-modelled by ellipse fitting to show that this provides a more representative fit as suggested in Inceoglu et al. (2014).

The correlation between measured CR intensities and modelled CR intensities for cycles 20–23 are presented in Table 4.5.

Table 4.5: Correlation coefficients of the linear regression and ellipse modelling of the hysteresis plots for solar cycles 20–23.

	Cycle 20		Cycle 21		Cycle 22		Cycle 23	
	Linear	Ellipse	Linear	Ellipse	Linear	Ellipse	Linear	Ellipse
McMurdo	0.867	-	0.664	0.946	0.964	-	0.846	0.852
Newark	0.888	-	0.700	0.964	0.955	-	0.857	0.874
South Pole	0.746	-	0.358	0.939	0.936	-	0.733	0.855
Thule	0.783	-	0.912	0.964	0.900	-	0.813	0.929

There is a consistent and good agreement between the measured and modelled CR intensities for even solar cycles modelled through linear regression, because the hysteresis loops are quite narrow as shown in Figure 4.7. These results support the findings of Inceoglu et al. (2014). Note that discrepancies in the correlation coefficients between this study and Inceoglu et al. (2014) are likely due to a number of reasons: Inceoglu et al. (2014) used data smoothing processes where in this study raw data are used; Inceoglu et al. (2014) made use of monthly mean data, whereas annual mean data was used in this study; Inceoglu et al. (2014) interpolated missing data, whereas gaps have been left untreated in this study.

The linear relations for odd solar cycles are less consistent in their agreement with observed CR intensities, with the correlation during cycle 21 as low as 0.34 for South Pole and as high as 0.91 for Thule. Across both of the odd cycles considered in this study linear regression is not as good a representation of the data as for even cycles. Figure 4.8 shows the wider hysteresis loops which is a characteristic of odd solar cycles and shows visually that a linear fit does not provide a good representation of the data.

In agreement with the results of Inceoglu et al. (2014), it can be seen from the results in Table 4.5 that the ellipse models provide estimates of the CR intensity which are in good agreement with the measured intensities due to the increased

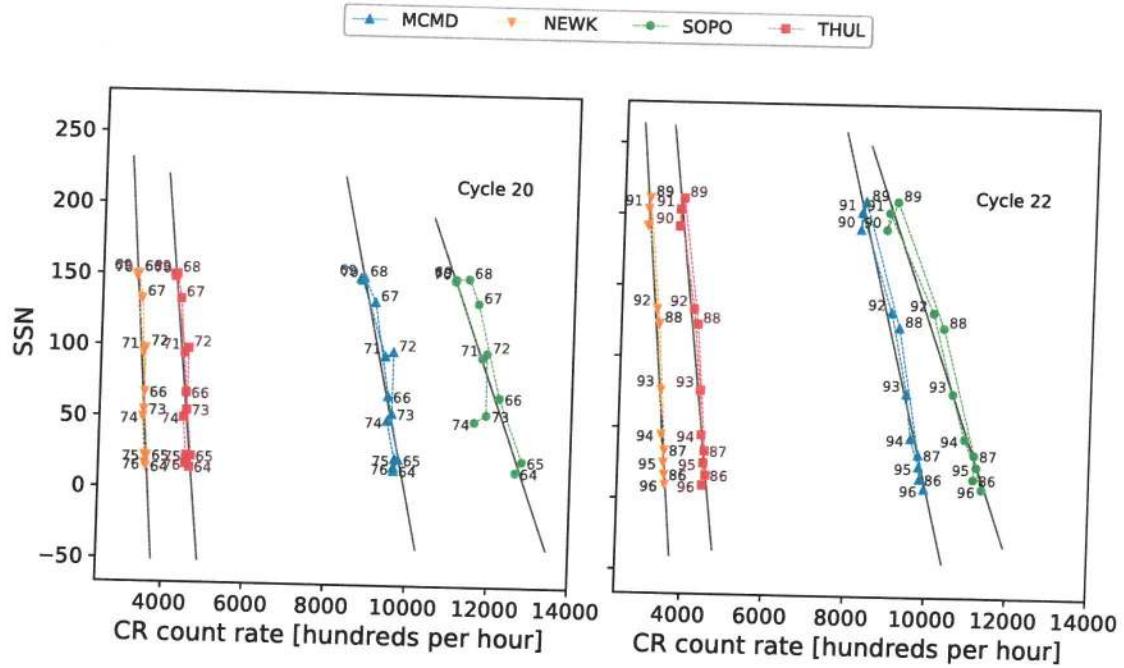


Figure 4.7: The hysteresis plots for even solar cycles 20 and 22 and the linear regression fit to the data.

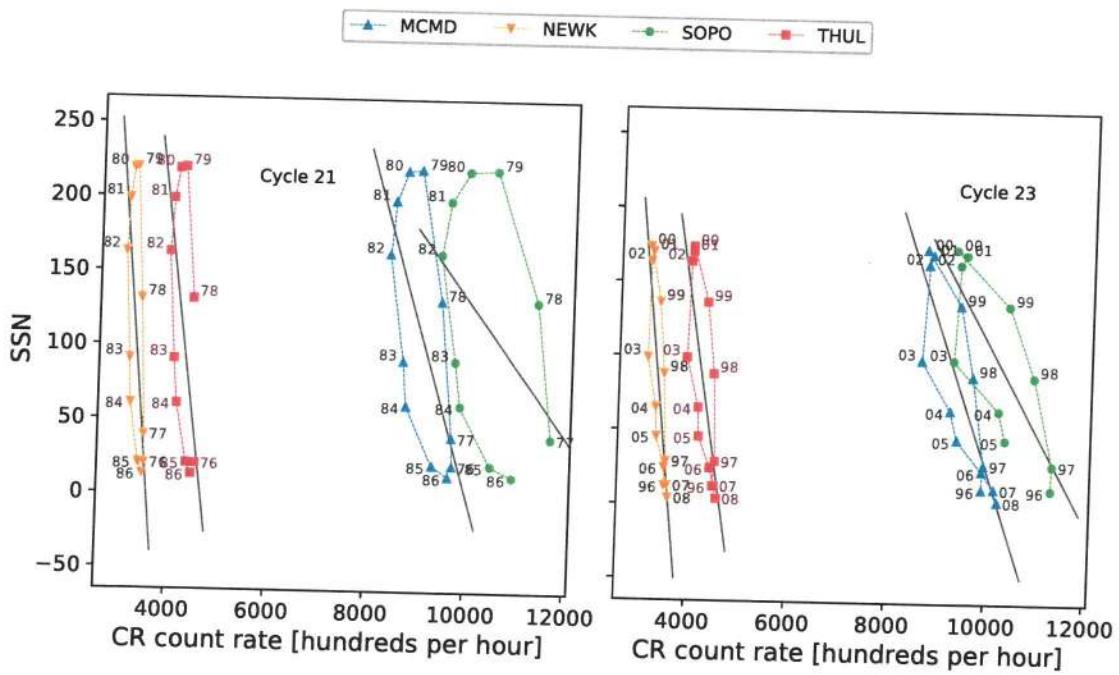


Figure 4.8: The hysteresis plots for odd solar cycles 21 and 23 and the linear regression fit to the data.

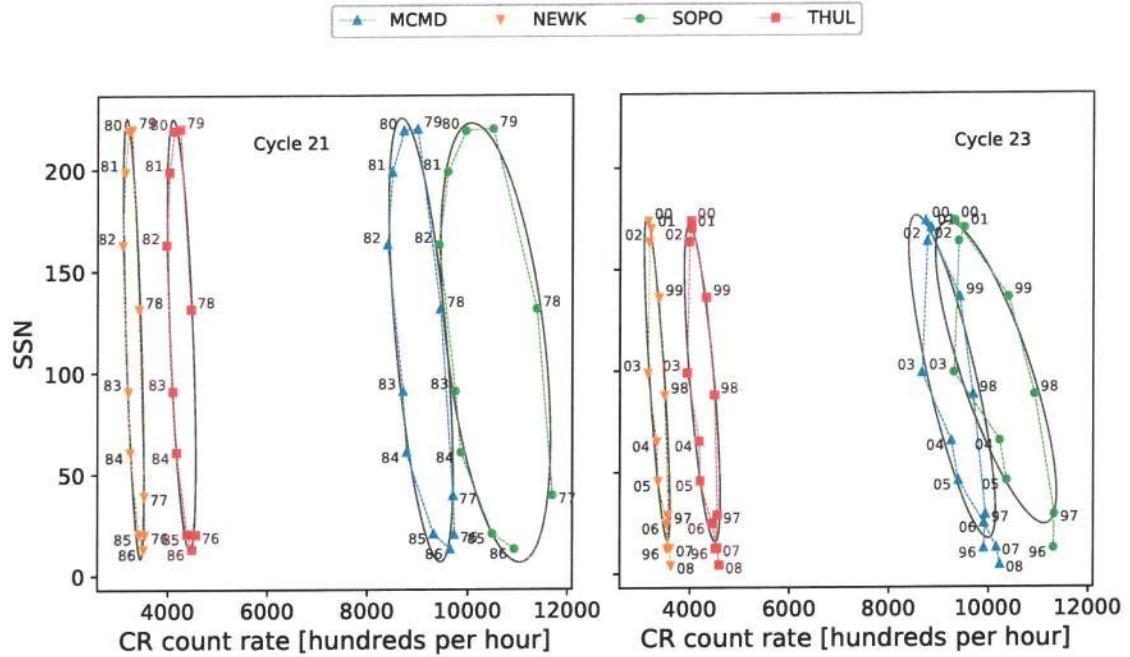


Figure 4.9: The hysteresis plots for odd solar cycles 21 and 23 and the ellipse fit to the data.

correlation coefficient for each station during cycle 21 and cycle 23; for SOPO during cycle 21 the increase in ρ is seen to be 0.58 proving the benefit of the ellipse model.

If cycle 24 follows the pattern between odd and even cycles, it is expected that the best fit will be provided by the linear model; however, it can be seen in Figure 4.2 that cycle 24 appears to display a wider hysteresis loop than the two preceding even-numbered cycles. Both the linear model and ellipse model were applied to the hysteresis plots for solar cycle 24 to determine which model would provide the better fit; the correlation between measured CR intensities and modelled CR intensities are presented in Table 4.6.

Table 4.6: Correlation coefficients of the linear regression and ellipse modelling of the hysteresis plots for solar cycle 24.

	Cycle 24	
	Linear	Ellipse
McMurdo	0.903	0.927
Newark	0.936	-
South Pole	0.883	-
Thule	0.873	0.936

(Handwritten note: Think you should be able to fit parameters for this.)
answered in Viva.

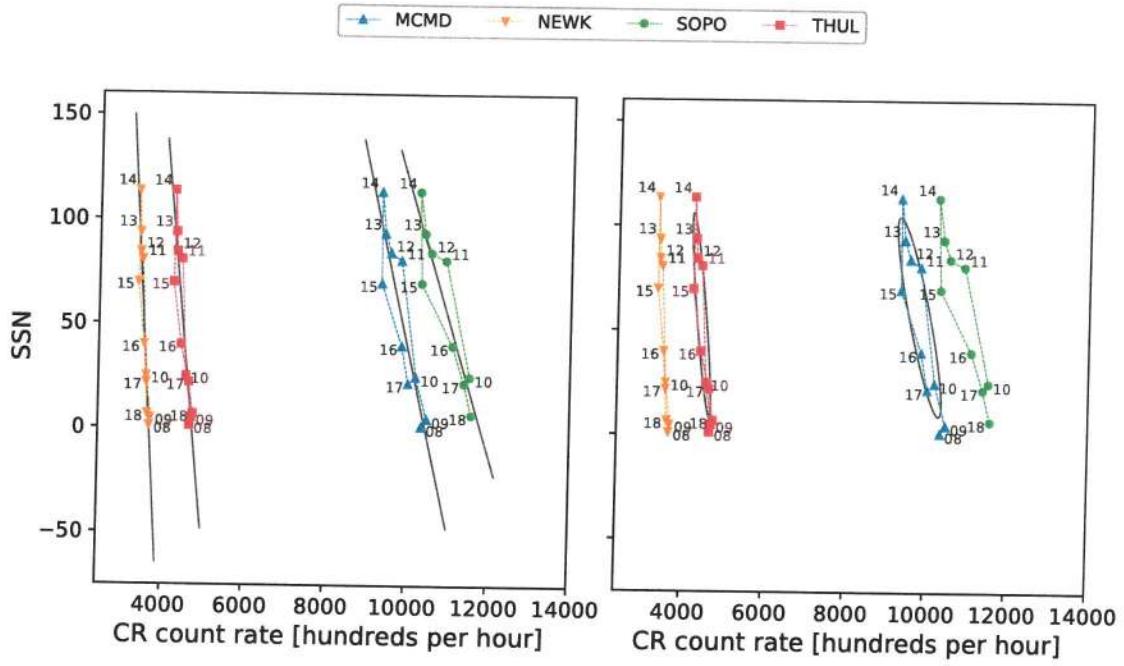


Figure 4.10: The hysteresis plot for solar cycle 24, and the linear regression fit to the data (left) and ellipse fit to the data (right).

The linear model for cycle 24 shows a good correlation between the observed and modelled CR intensities providing evidence to suggest that cycle 24 follows the two preceding even-numbered cycles. The ellipse model does however improve the relation between the observed and modelled CR intensities for 2 out of the 4 stations: McMurdo and Thule. For South Pole and Newark the ellipse model was not able to provide a fit at all, which is believed to be due to the Newark data points crossing where the semi-major axis would be defined for the ellipse model causing the calculation of the semi-major and semi-minor axes to return as not a number, and the South Pole has data missing at the beginning of cycle 24.

The results for solar cycle 24 do not provide a conclusive answer as to whether cycle 24 behaves like past even solar cycles or odd cycles from this data set alone; however, the ellipse model does not provide as significant an improvement for the two modelled NM stations as for odd cycles. The small improvement in the ellipse fit is likely due again to the effects of the extended declining phase of cycle 23 and the unusually low activity of cycle 24.

*answered
in
Viva*

*Is this looks like
works by
fit an
ellipse?*

We repeated the analysis for the additional NM stations that featured in this study. Again, the linear model provided a good fit to the data however the ellipse model was not able to provide a fit; favouring the conclusion that cycle 24 is best represented by a simple linear model as was true for the preceding even-numbered cycles.

Despite cycle 24 having not yet declined to a minimum, it is clear from the observations shown in the hysteresis plots that further data in cycle 24 is unlikely to broaden the loop any further. The hysteresis loop begins to tighten up after 2016 following the broadening between 2014–2016; hence it appears unlikely that by the end of cycle 24 further observations will support the ellipse model and instead will favour the linear fit to the data.

4.5 Conclusions

As cosmic rays are modulated by the heliosphere during the 11-year solar activity cycle, and this effect has been studied for previous solar cycles, the principal aim of this study was to investigate the nature of GCRs during the current activity cycle 24 as it draws to a minimum.

In this study we presented a time-lag analysis between GCR intensity and SSN which showed that cycle 24 has a longer lag (2–4 months) than the preceding even-numbered solar activity cycles (typically 0–1 months); however, its lag is not as large as preceding odd-numbered cycles, and cycle 24 follows the trend of a short or near-zero lag for even-numbered cycles. We suggest here that the cause of the extended lag in cycle 24 compared to previous even-numbered cycles is likely due to the deep, extended minimum between cycle 23 and 24, and the low maximum activity of cycle 24 (Broomhall, 2017).

It has been previously shown in the literature that there is a striking difference in the shape of the plot of SSN and GCR intensity between odd-numbered and even-numbered solar cycles. Due to the difference in the shape of the hysteresis plots

for odd-numbered and even-numbered cycles, we have modelled the hysteresis plots using both a simple linear model and an ellipse model. The results of this study tend to support that cycle 24 follows the same trend as preceding even-numbered cycles and is best represented by a straight line rather than an ellipse, such is the case for odd-numbered activity cycles.

We emphasise that although cycle 24 has not yet ended, the shape of the hysteresis plots suggest that we are now past the main broadening region and the inclusion of further data for cycle 24 will very likely only support the linear model. This study will continue to follow the evolution of the cycle 24 until the onset of cycle 25, in around 2019–2021 (Howe et al., 2018; Upton & Hathaway, 2018; Pesnell & Schatten, 2018), when an update on the final results of cycle 24 should be provided.

4.6 Comparison with HiSPARC

The analysis was repeated using data from HiSPARC station 501, over the period between 2008 – 2019, to investigate whether the HiSPARC (HS) network was capable of monitoring the solar cycle variation in the muon count rate. HiSPARC station 501 was the best candidate for this analysis due to its long operational lifetime, in comparison to other HS stations, which spans most of cycle 24, and due to the fact that this station is used as the *gold standard* HS station.

This was further motivated by the observations of Fan & Velthuis (2018), in which they stated a confirmation of the accepted anti-correlation between solar activity and CRs after analysing the count rate of the HS 501 station, up to 2016/17, comparing the count rate with the number of solar flares and the SSN during the rising phase of cycle 24. We also wanted to determine whether the results of the analysis for the behaviour of GCRs during cycle 24, using HS data, agreed with those acquired above, using NM data.

A plot of the HS 501 data over this period is given in Figure 4.11, which shows a comparison between the HS count rate and SSN. By-eye, no clear anti-correlation

Cycle 24 ended
in Dec 2019. I know
this is from
your paper,
but this
could be
updated.

is displayed between the HS count rate and the SSN.

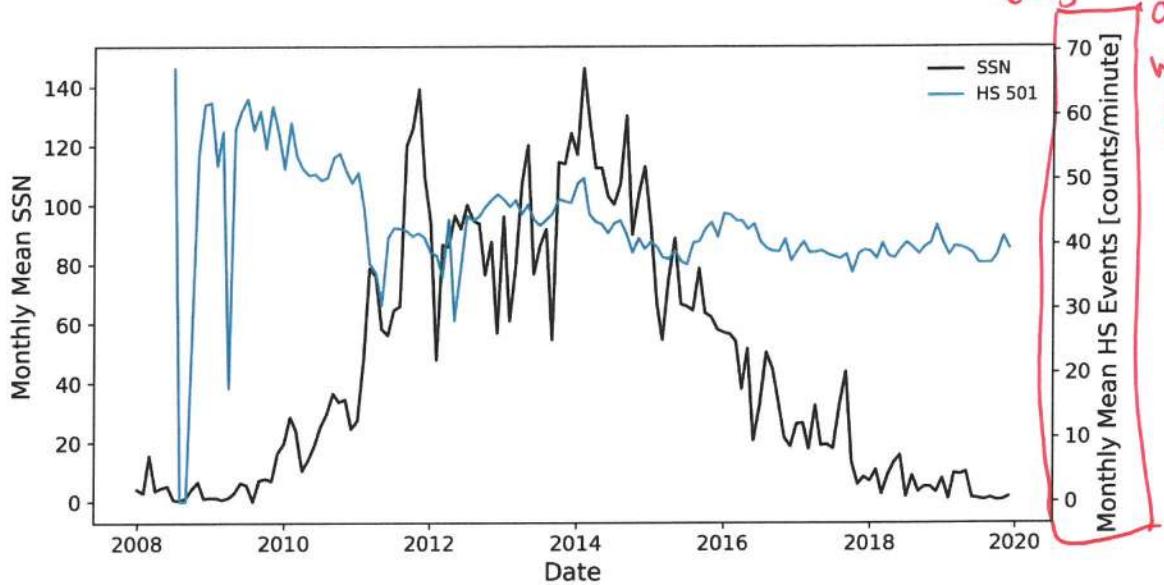


Figure 4.11: The monthly-mean CR-induced muon count rate recorded by HiSPARC (blue), and the SSN (black) between 2008 – 2019.

The spearman correlation coefficient between the HS count rate and the SSN was calculated between the two data sets, without performing the time-lag analysis, to be, $\rho = 0.078$. It was quantitatively concluded therefore that there exists no correlation between the SSN and the HS muon count rate, confirming what was suspected by-eye.

This lack of anti-correlation is a consequence of the operation of the HS stations. In many stations, there is a propensity to regularly change the detector voltages and threshold settings, which limits the capability of HS to observe the solar cycle modulation of GCR intensity. The HiSPARC stations are typically set up to achieve a singles rate of around 100 Hz above the higher threshold limit, resulting in an events rate of around 1 Hz for the station (Fokkema, 2012), in order to achieve their scientific goals. The Photo Multiplier Tube (PMT) operating High Voltages (HVs) are often adjusted when the count rate strays away from this nominal value, which mitigates the ability of HiSPARC to reliably observe the solar cycle variation.

Modifications to the PMT operating HV of each detector in HS 501 are illustrated in Figure 4.12, and shows the scale of the adjustments, with the voltages varying

by as much as ~ 80 V. The effect of varying the PMT voltage is dependent on the individual PMT, but a change of 50–100 V will have an effect on the count rate, therefore it is likely to be a root cause for not observing the solar cycle modulation.

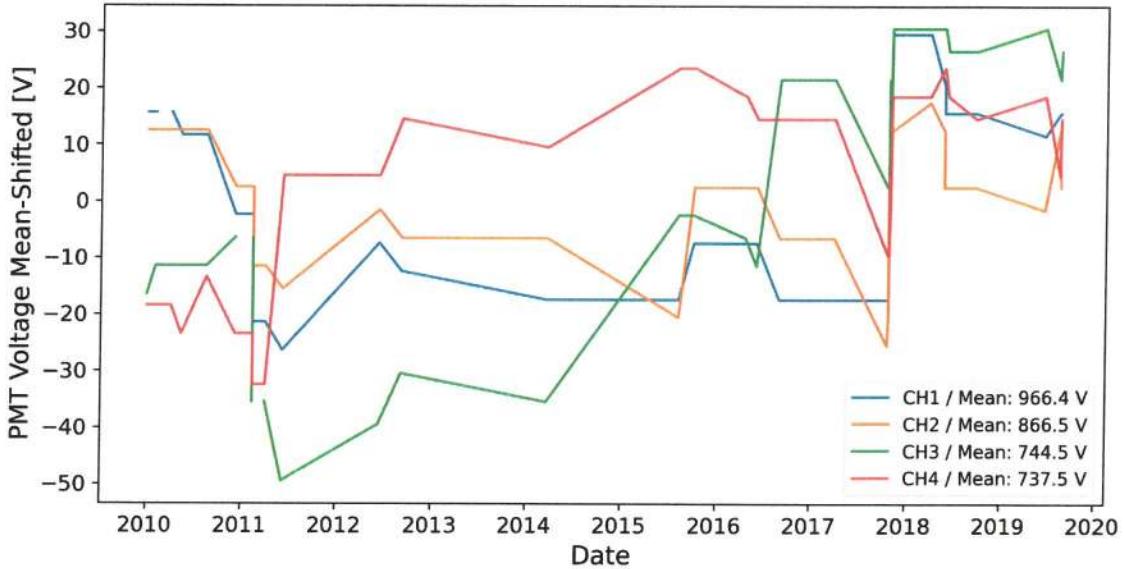


Figure 4.12: The evolution of the mean-shifted PMT operating voltages for each detector within HiSPARC station 501 between 2010 – 2020.

The claim that the solar cycle variation was observed in the count rate of HS 501 by Fan & Velthuis (2018) was highly likely based on their limited data set, spanning not enough of solar cycle 24. One could be convinced from the count rate during the rising phase of cycle 24 that the cycle variation is observed, but clearly beyond 2016 this relationship does not continue. It will be extremely interesting to re-visit this analysis of the HiSPARC data at the end of solar cycle 25, and particularly to repeat this analysis with the data from the new station configuration, HiSPARC station 14008.

5 A Frequency Domain Investigation on the Morphology of the Solar Mean Magnetic Field

Most of the content in this chapter has been restructured for publication in the journal Monthly Notices of the Royal Astronomical Society (Ross et al., 2021). I was first author on this journal article and conducted the entirety of the work in the investigation.

5.1 Introduction

The Sun has a complicated magnetic field structure; many features of the Sun and proxies for the solar activity are related to the evolution of the Sun’s magnetic field (Wu et al., 2018).

The Solar Mean Magnetic Field (SMMF) is a surprising, non-zero measurement of the imbalance of opposite magnetic flux polarities observed on the full, visible disc of the Sun (Svalgaard et al., 1975), and is defined as the mean Line Of Sight (LOS) magnetic field when observing the Sun-as-a-star (Scherrer et al., 1977a,b; Garca et al., 1999). In the literature the SMMF is also sometimes referred to as the General Magnetic Field (GMF) (Severny, 1971) or the Mean Magnetic Field (MMF) (Kotov, 2008) of the Sun.

Observations of the SMMF have typically been made by measuring the Zeeman

by as much as ~ 80 V. The effect of varying the PMT voltage is dependent on the individual PMT, but a change of 50–100 V will have an effect on the count rate, therefore it is likely to be a root cause for not observing the solar cycle modulation.

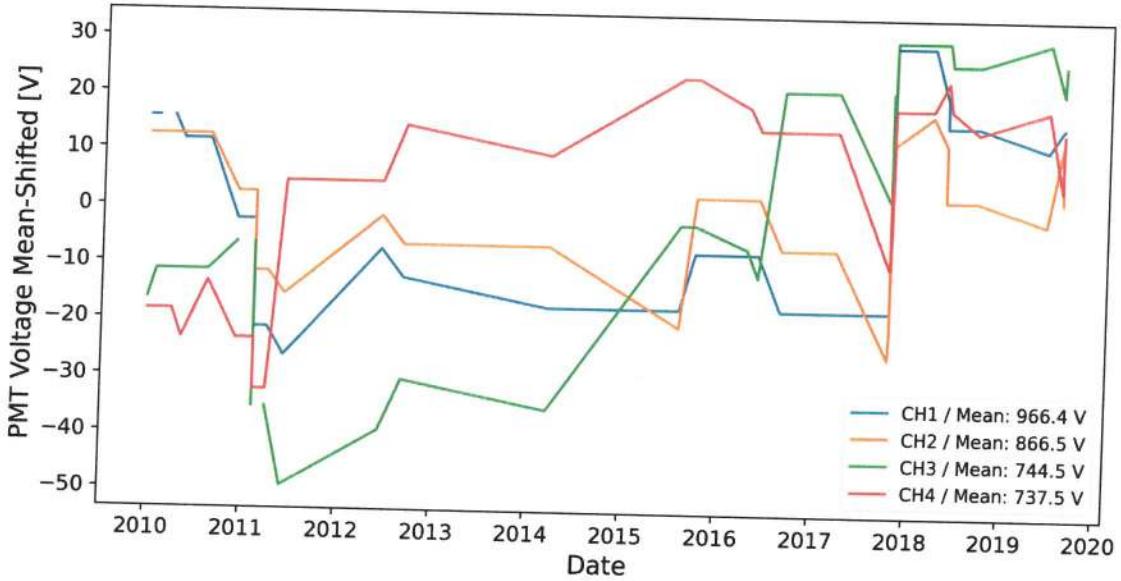


Figure 4.12: The evolution of the mean-shifted PMT operating voltages for each detector within HiSPARC station 501 between 2010 – 2020.

The claim that the solar cycle variation was observed in the count rate of HS 501 by Fan & Velthuis (2018) was highly likely based on their limited data set, spanning not enough of solar cycle 24. One could be convinced from the count rate during the rising phase of cycle 24 that the cycle variation is observed, but clearly beyond 2016 this relationship does not continue. It will be extremely interesting to re-visit this analysis of the HiSPARC data at the end of solar cycle 25, and particularly to repeat this analysis with the data from the new station configuration, HiSPARC station 14008.

5 A Frequency Domain Investigation on the Morphology of the Solar Mean Magnetic Field

Most of the content in this chapter has been restructured for publication in the journal Monthly Notices of the Royal Astronomical Society (Ross et al., 2021). I was first author on this journal article and conducted the entirety of the work in the investigation.

5.1 Introduction

The Sun has a complicated magnetic field structure; many features of the Sun and proxies for the solar activity are related to the evolution of the Sun’s magnetic field (Wu et al., 2018).

The Solar Mean Magnetic Field (SMMF) is a surprising, non-zero measurement of the imbalance of opposite magnetic flux polarities observed on the full, visible disc of the Sun (Svalgaard et al., 1975), and is defined as the mean Line Of Sight (LOS) magnetic field when observing the Sun-as-a-star (Scherrer et al., 1977a,b; Garca et al., 1999). In the literature the SMMF is also sometimes referred to as the General Magnetic Field (GMF) (Severny, 1971) or the Mean Magnetic Field (MMF) (Kotov, 2008) of the Sun.

Observations of the SMMF have typically been made by measuring the Zeeman

splitting of spectral lines using a ground-based Babcock-type magnetograph (Scherrer et al., 1977a), although more recently the SMMF has been calculated from full-disc LOS magnetograms taken from space-borne telescopes such as the Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI), in order to understand the morphology of the SMMF (Kutsenko et al., 2017; Bose & Nagaraju, 2018). It is understood that the strength of the SMMF may vary depending on the spectral line used to measure the SMMF (Kotov, 2008, 2012); however, the SMMF varies slowly with the solar activity cycle, with an amplitude on the order of a Gauss during solar maximum and a tenth of a Gauss during solar minimum (Plachinda et al., 2011). In addition, the SMMF displays a strong, quasi-coherent rotational signal, which must arise from inhomogeneities on the solar disc with lifetimes of several rotations (Chaplin et al., 2003; Xie et al., 2017).

Despite existing literature on SMMF observations spanning several decades, ultimately, the origin of the SMMF remains an open question in solar physics. The principle component of the SMMF is commonly assumed to be weak, large-scale magnetic flux, distributed over vast areas over the entire solar disc, rather than from more concentrated regions such as Active Regions (ARs) or sunspots (Severny, 1971; Scherrer et al., 1977a; Xiang & Qu, 2016). However, conversely, Scherrer et al. (1972) found that the SMMF was most highly correlated with flux from only the inner-most one quarter, by area, of the solar disc, which is more sensitive to active latitudes.

In recent literature, Bose & Nagaraju (2018) provided a novel approach to understanding the SMMF whereby they decomposed the SMMF through feature identification and pixel-by-pixel analysis of full-disc magnetograms. Bose & Nagaraju (2018) concluded that: (i) the observed variability in the SMMF lies in the polarity imbalance of large-scale magnetic field structures on the visible surface of the Sun, (ii) the correlation between the flux from sunspots and the SMMF is statistically insignificant, (iii) and more critically that the background flux dominates the

SMMF, accounting for around 89% of the variation in the SMMF. However, there still remained a strong manifestation of the rotation in the background magnetic field presented by Bose & Nagaraju (2018). This is indicative of inhomogeneous magnetic features with lifetimes on the order of several solar rotations rather than the shorter-lived, weaker fields usually associated with the large-scale background. It therefore raises the question of whether their technique assigned flux from Magnetic Flux Concentrations (MFCs) or ARs to the background.

In order to identify the contours of specific features Bose & Nagaraju (2018) used an adaptive thresholding technique on various Solar Dynamics Observatory Atmospheric Imaging Assembly (SDO/AIA) images of the solar disc to create binary masks for different types of features. These masks were then applied to scaled SDO/HMI magnetograms in order to separate the features on the disc and their contribution to the SMMF. Upon a closer inspection of the example magnetogram in Figure 2 of the paper, with over-plotted contours of identified features from SDO/AIA images, there are clearly regions of strong MFCs in the local vicinity of, and connected to, the identified features that lie outside their contour lines. Therefore these were allocated to the background magnetic field, rather than attributed to the specific features. It seems an obvious statement to suggest that SDO/AIA optical counterparts of the magnetograms will not exactly align with the observed magnetic flux in the magnetograms. We expect that the magnetic field will manifest itself differently in the optical observations and the magnetograms, which leads one to believe that the background component in this study could mistakenly contain flux from some of the identified features.

in you
include this
here.

if the fig
is in your
thesis it will
be easy
to see this!

In particular, in their example plot, the contours for sunspots typically only cover the umbra, not accounting for the surrounding penumbra. It is not clear whether this will have a large effect, but certainly we expect that some strong magnetic flux associated with sunspots has been attributed to the background, or other nearby features. One other note; there was a treatment of plages in this

work, from additional chromospheric observations, but a separate, specific handling of faculae in the photosphere was absent. Incorporating this could have contributed to the completeness of the study. Furthermore, a decomposition of the identified background component into regimes of strong and weak field would have provided more clarity on the exact morphology of the SMMF, and would have likely provided evidence to conclude whether flux from AR features were, in fact, incorporated into the background.

Despite these findings, it is known that the strength of the SMMF is weaker during solar minimum, when there are fewer ARs, and stronger during solar maximum, when there are more ARs. This is suggestive that the evolution of ARs has relevance for the evolution of the SMMF.

There is a contrasting view in the literature which suggests AR flux dominates the SMMF. [Kutsenko et al. \(2017\)](#) claim that a large component of the SMMF may be explained by strong and intermediate flux regions. These regions are associated with ARs; using a thresholding technique they showed between 65% to 95% of the SMMF could be attributed to strong and intermediate flux, while the fraction of the occupied area varied between 2% to 6% of the disc area, depending on the chosen threshold for separating weak and strong flux. This finding suggests that strong, long-lived, inhomogeneous MFCs produce the strong rotation signal in the SMMF. Potential sources could be sunspots, plages, faculae, etc. and [Kutsenko et al. \(2017\)](#) stated that there is an entanglement of strong flux (typically associated with ARs) and intermediate flux (typically associated with network fields and remains of decayed ARs). Disentangling the flux would have provided a more accurate analysis of the SMMF, owing to a clearer picture of the main contributor to the SMMF.

The Sun's dynamo and hence magnetic field is directly coupled to the solar rotation. The Sun exhibits latitude-dependent and depth-dependent differential rotation with a sidereal, equatorial period of around 25 days ([Howe, 2009](#)). To Earth-based observers, the synodic rotation of the Sun is observed at around 27 days,

and the SMMF displays a dominant periodicity of around 27 days due to the solar rotation (Chaplin et al., 2003; Xie et al., 2017; Bose & Nagaraju, 2018). It was also reported by Xie et al. (2017) that the surface differential rotation was observed in the SMMF with measured rotational periods of 28.28 ± 0.67 days and 27.32 ± 0.64 days for the rising and declining phases, respectively, of all of the solar cycles in their considered time-frame.

On the other hand, Xiang & Qu (2016) utilised ensemble Empirical Mode Decomposition (EEMD) analysis to extract rotational modes in the SMMF and found two rotation periods which are derived from different strengths of magnetic flux elements. They found that a rotation period of 26.6 days was related to a weaker magnetic flux element within the SMMF, while for stronger magnetic flux elements in the SMMF, the measured rotation period was 28.5 days.

Ultimately, to date, our understanding of the SMMF and its origin remains open to question.

5.2 Aims

In this work an investigation of high-cadence (sub-minute) observations of the SMMF, made by Birmingham Solar Oscillations Network (BiSON) (Chaplin et al., 1996, 2005; Hale et al., 2016), was performed. The aim of the investigation was to understand the morphology of the SMMF.

This work provides a frequency domain analysis of the SMMF data, where a model was built and fit to the power spectrum of the SMMF which allowed us to understand the characteristics of its source(s).

The Rotationally Modulated (RM) signal in the SMMF was clearly observed as several low-frequency peaks in the power spectrum. In addition, the use of the high-cadence data was especially crucial for inferences on components of the SMMF with periods of less than a day at higher frequencies in the power spectrum, with the intention to determine whether the background magnetic field exhibited a stochas-

tically excited component, which evolved on short timescales.

5.3 Data

5.3.1 Summary of the Data Set

Chaplin et al. (2003) provided the first examination of the SMMF using data from BiSON, and the work presented in this paper is a continuation of that study.

BiSON is a six-station, ground-based, global network of telescopes continuously monitoring the Sun, which principally makes precise measurements of the LOS velocity of the photosphere due to solar *p* mode oscillations (Hale et al., 2016). Through the use of polarising optics and additional electronics, the BiSON spectrometers can measure both the disc-averaged LOS velocity and magnetic field in the photosphere (Chaplin et al., 2003), however, not all BiSON sites measure the SMMF.

In this study we focus on the data collected by the Sutherland node, in South Africa, which was also used by Chaplin et al. (2003). Data are sampled on a 40-second cadence, and the SMMF data collected by the Sutherland station pertains to the epochs from 01/1992 – 12/2012 (i.e. covering 7643 days). Over this period, the fill (i.e. duty cycle) of solar observations is low because of the combination of: (a) the fill due to using only a single site with a 40-second cadence, (b) the duty cycle needed in a given day for a ‘good measure’ of the Sun. Over this period, the average duty cycle of the 40-second data is \sim 15.6%. If instead we take a daily average of the BiSON SMMF, the average duty cycle is \sim 55.2%. This gives a higher duty cycle but a lower Nyquist frequency. Because of the much lower Nyquist frequency, modelling the background power spectral density is more challenging; therefore we use the 40-second cadence data in this work.

As a comparison to the BiSON data, SMMF observations were also acquired by the Wilcox Solar Observatory (WSO) (<http://wso.stanford.edu/>) (Scherrer et al., 1977b). The WSO SMMF data are sampled daily from 16/05/1975 – present day, but for comparison with the BiSON SMMF, we only used data over the same

temporal range.

The WSO also measures the LOS SMMF using a Babcock-type magnetograph, which allows the measurement of the amount and sense of circular polarisation in the wings in an absorption line (Scherrer et al., 1977a). WSO uses two absorption lines for the measurement of the magnetic field: the Fe I at 5250 Å(λ 5250) is used for measurement of the field, and Fe I at 5124 Å(λ 5124) is used to determine the instrument's zero offset, as this line is magnetically insensitive. Scherrer et al. (1977a) describes that a single, complete observation takes \sim 20 minutes, consisting of four 3-minute integrations. The data provided by WSO are a daily weighted mean of the observations, where the weighting used is the statistical uncertainty and the magnitude of the zero offset measured with the λ 5124 line.

5.3.2 Obtaining the SMMF from BiSON

There is no catalogued BiSON SMMF data-set, so it was necessary to compute the SMMF from the available BiSON data. To acquire the SMMF from BiSON data, the method as described by Chaplin et al. (2003) was adopted; here we discuss the key aspects.

Each BiSON site employs a Resonant Scattering Spectrometer (RSS) to measure the Doppler shift of the $^2\text{S}_{1/2} \rightarrow ^2\text{P}_{1/2}$ line (D1 line) of potassium, at \sim 770 nm (Brookes et al., 1978). A potassium vapour cell placed within a longitudinal magnetic field Zeeman splits the laboratory line into the two allowed D1 transitions (Lund et al., 2017). The intensity of the longer wavelength (red; I_R) and shorter wavelength (blue; I_B) components of the line may be measured by the RSS almost simultaneously, by using polarising optics to switch between the red and blue wings of the line, to form the ratio given by equation (5.1), which is used as a proxy for the Doppler shift from the LOS velocity of the photosphere (see: Brookes et al., 1976,

1978; Elsworth et al., 1995a; Chaplin et al., 2003; Lund et al., 2017):

$$\mathcal{R} = \frac{I_B - I_R}{I_B + I_R}. \quad (5.1)$$

Photospheric magnetic fields Zeeman split the Fraunhofer line and the Zeeman-split components have opposite senses of circular polarisation (Chaplin et al., 2003). Additional polarising optics are used in the RSS to manipulate the sense of circular polarisation (either + or -) that is passed through the instrument. The ratio \mathcal{R}_+ or \mathcal{R}_- is formed, and the ratios \mathcal{R}_\pm would be equal if there was no magnetic field present.

The observed ratio (\mathcal{R}_\pm) may be decomposed as:

$$\mathcal{R}_\pm = \mathcal{R}_{\text{orb}} + \mathcal{R}_{\text{spin}} + \mathcal{R}_{\text{grs}} + \delta r_{\text{osc}}(t) \pm \delta r_B(t), \quad (5.2)$$

where \mathcal{R}_{orb} is due to the radial component of the Earth's orbital velocity around the Sun, $\mathcal{R}_{\text{spin}}$ is due to the component towards the Sun of the Earth's diurnal rotation about its spin axis as a function of latitude and time, \mathcal{R}_{grs} is from the gravitational red-shift of the solar line due to the Sun's mass (Elsworth et al., 1995b; Dumbill, 1999), $\delta r_{\text{osc}}(t)$ is due to the LOS velocity of p mode oscillations, and $\delta r_B(t)$ is due to the magnetic field (\pm denotes the polarity of the Zeeman-split line that is being observed) (Dumbill, 1999).

The effect of the magnetic field on the ratio is shown in Figure 5.1, and it is clear to see from equation (5.3) that the difference between the opposite magnetic field ratios is twice the magnetic ratio residual, i.e.:

$$\mathcal{R}_+ - \mathcal{R}_- = 2 \delta r_B(t). \quad (5.3)$$

The BiSON RSS is measuring the velocity variation on the solar disc, and therefore a calibration from the ratio to a velocity is necessary. One method of velocity

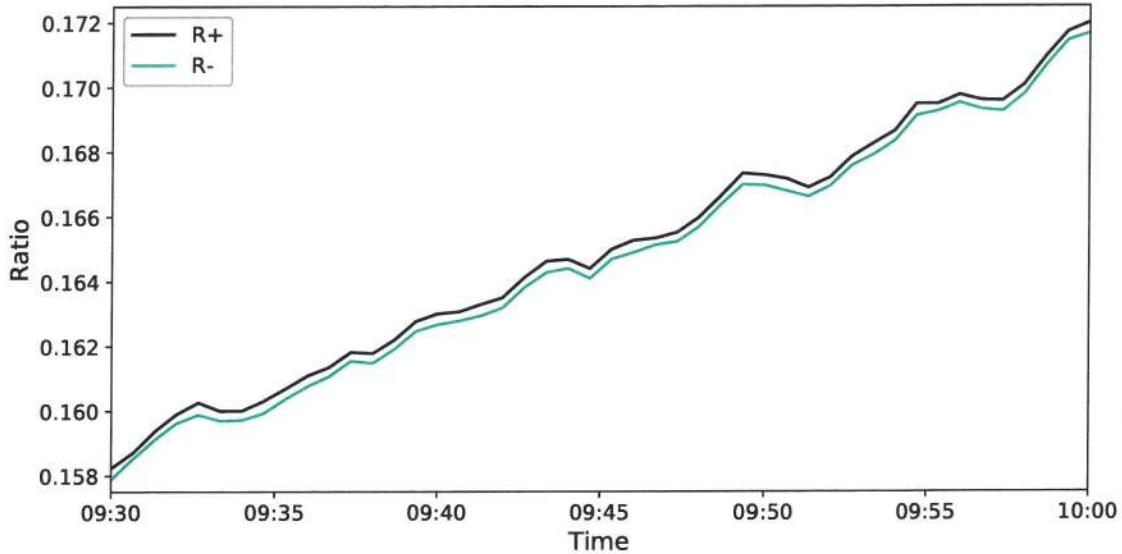


Figure 5.1: An example of the BiSON ratios data over a 30-minute period. The separation between the two ratios is due to the solar mean magnetic field. Other excursions in the individual ratios are due to the other effects measured by the RSS.

calibration is achieved by first fitting the daily observed ratio, averaged over both magnetic polarities, to a 2nd- or 3rd-order polynomial as a function of velocity, as discussed by Elsworth et al. (1995b). Here we chose to fit the ratio in terms of velocity, $\mathcal{R}_{\text{calc}}(u)$, i.e.:

$$\mathcal{R}_{\text{calc}}(u) = \sum_n \mathcal{R}_n u^n, \quad (5.4)$$

where:

$$u = v_{\text{orb}} + v_{\text{spin}}, \quad (5.5)$$

and v_{orb} is the velocity component related to the ratio, \mathcal{R}_{orb} ; v_{spin} is related to the ratio, $\mathcal{R}_{\text{spin}}$; and n is the polynomial order.

It is possible to see that through the removal of $\mathcal{R}_{\text{calc}}(u)$ from the observed ratios, one is left with the ratio residuals of the p mode oscillations and the magnetic field, i.e.:

$$\mathcal{R}_{\pm} - \mathcal{R}_{\text{calc}}(u) = \delta r_{\text{osc}}(t) \pm \delta r_{\text{B}}(t). \quad (5.6)$$

Furthermore, conversion from ratio residuals into velocity residuals uses the calibration given by:

$$\delta v(t) = \left(\frac{d\mathcal{R}_{calc}}{dV} \right)^{-1} \delta r(t). \quad (5.7)$$

In order to finally obtain the SMMF in magnetic field units, one must combine equation (5.3) and equation (5.7) with the conversion factor in equation (5.9) (Dumbill, 1999), and the entire procedure can be simplified into:

$$B(t) = \frac{1}{2} \left(\frac{d\mathcal{R}_{calc}}{dV} \right)^{-1} \frac{(\mathcal{R}_+ - \mathcal{R}_-)}{K_B}, \quad (5.8)$$

where:

$$K_B = \frac{8}{3} \frac{\mu_B}{h} \frac{c}{\nu} \approx 2.89 \text{ ms}^{-1} \text{ G}^{-1}, \quad (5.9)$$

and μ_B is the Bohr magneton, h is Planck's constant, c is the speed of light, and ν is the frequency of the photons.

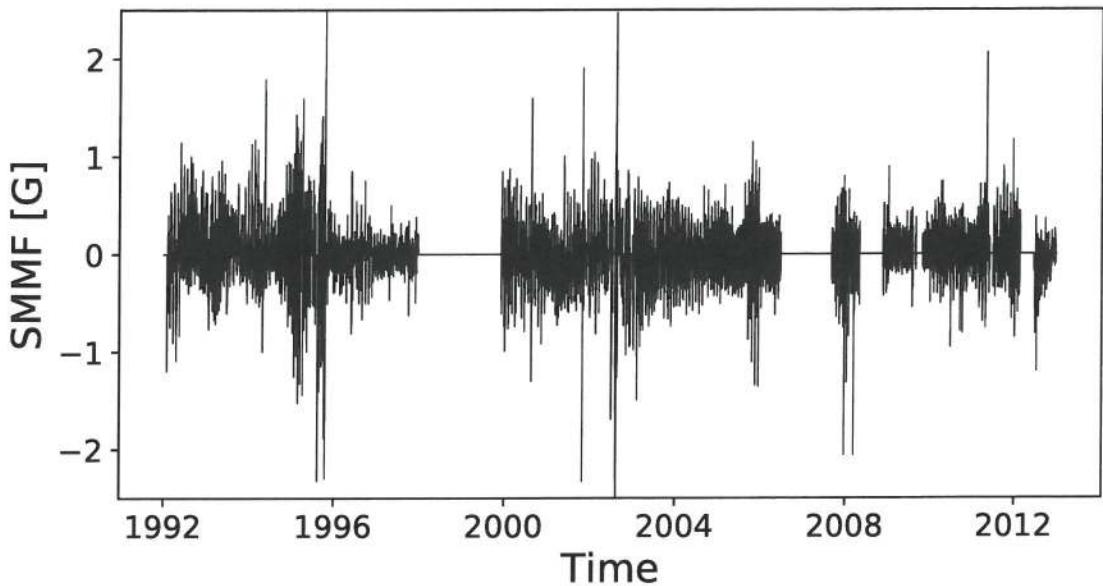
Through the application of this methodology, one acquires the SMMF as shown in Figure (5.2a). The power spectrum of the full, 7643-day Sutherland data set is shown in Figure (5.2b). The power spectrum shows a clear set of strong peaks at low frequency, which are due to the persistent rotation signal in the SMMF. The largest peak is the fundamental rotation frequency, and the following peaks are its harmonics.

5.4 Methodology

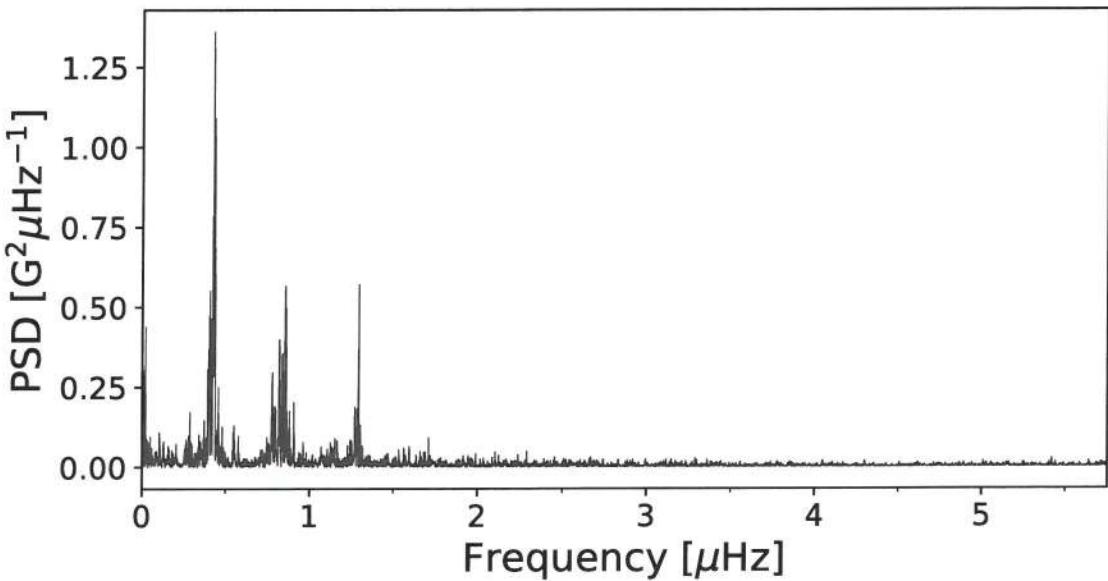
5.4.1 Identifying Features in the SMMF Power Spectrum

As we have 40-second cadence observations of the SMMF, we were able to investigate the power spectrum up to a Nyquist frequency of 12500 μHz . There are a number of features within the full power spectrum, shown in Figure 5.3.

The peaks between 0.1 – 2.0 μHz are a manifestation of a persistent rotational signal in the SMMF. The distinct set of peaks indicates the existence of a long-lived, inhomogeneous, RM source. The SMMF signal exhibits a quasi-coherent



(a) BiSON SMMF 40-second cadence time series



(b) Power spectral density of the BiSON SMMF

Figure 5.2: (a) 40-second cadence observations of the SMMF from the Sutherland BiSON station between 1992 and 2012. The sense of the field was chosen to match the [Chaplin et al. \(2003\)](#) and the WSO observations, where positive is for a field pointing outwards from the Sun. (b) Power spectrum of the SMMF on a 40-second cadence truncated to $10\mu\text{Hz}$, however, the Nyquist frequency is $12500\mu\text{Hz}$.

behaviour in the time domain, and based on the comparatively short timescales for the emergence of magnetic features compared to their slow decay (i.e. hours–days compared to weeks–months) ([Zwaan, 1981](#); [Harvey & Zwaan, 1993](#); [Hathaway &](#)

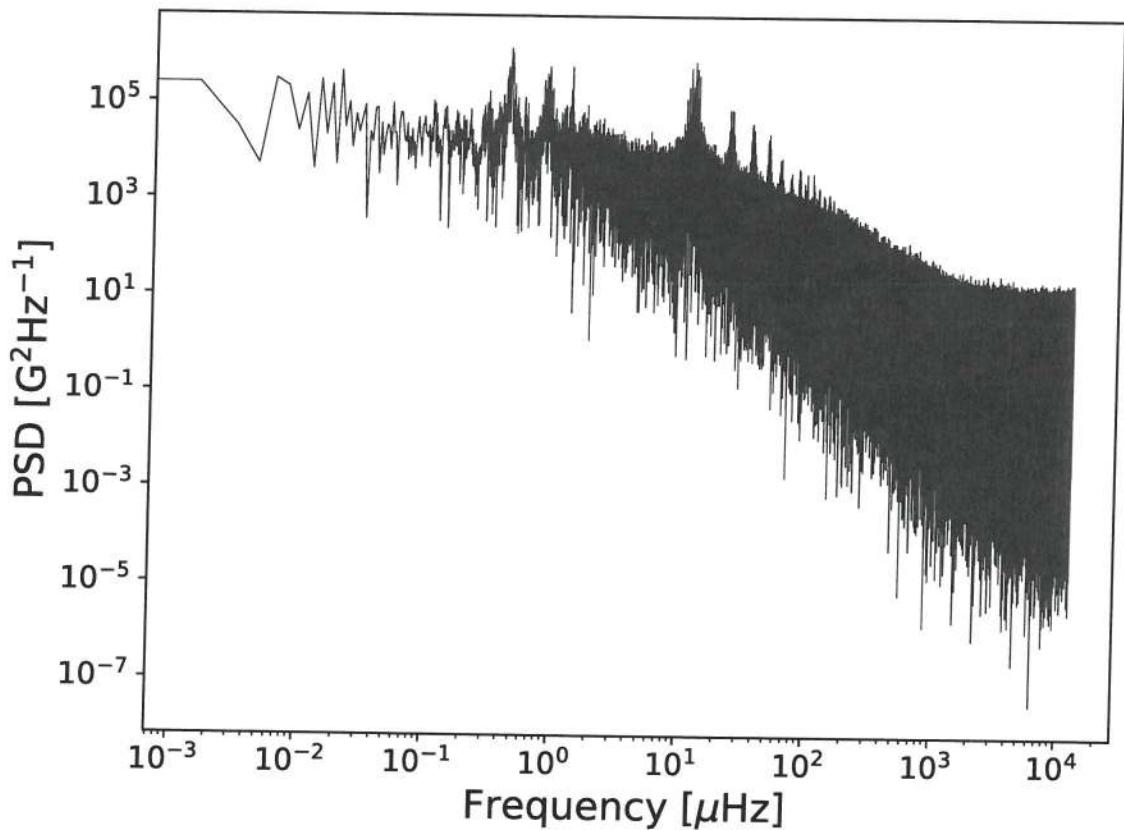


Figure 5.3: Power spectrum of 40-second cadence SMMF from the Sutherland BiSON station observed between 1992 – 2012 on a logarithmic scale up to the Nyquist frequency.

Choudhary, 2008; Dacie et al., 2016), we assume the evolution of the RM component with time is a sudden appearance and a long, exponential decay.

The high Nyquist frequency was critical in uncovering a red-noise-like component in the power spectrum. This component could arise from continuously evolving, short-lived regions of magnetic field linked to magneto-convection, akin to a random walk, which we will dub the Stochastic Background (SB) component. Analogous to the SB, is the granulation signal observed in the Doppler-velocity measurements of the solar surface (Basu & Chaplin, 2017). It is indeed possible that this is not a real feature in the data and it could also be present due to the high-cadence and low-fill of the observations.

In addition, at low-frequency there is power associated with instrumental noise and solar activity, and at very-high frequency shot-noise is captured which sets the

lower limit in power in the spectrum.

There are also side-band features in the power spectrum at multiples of 1/day \sim 11.57 μ Hz. The side-bands are a well-known phenomena in ground-based astronomical observations. They arise from gaps in the data which are a consequence of making single-site, ground-based observations of the Sun.

The duty cycle of the 40-second BiSON observations is very low, at around \sim 15%, therefore it was important to take into consideration the effect that gaps in the data have on the power spectrum. Gaps in the data cause an aliasing of power from actual signal frequencies spread to other frequencies in the spectrum, and the nature of the aliasing depends on the properties of the window function of the observations. It is also possible that the SB component is an effect of power aliasing due to the low duty cycle of the data. Hence, before modelling the power spectrum, the window function was well-characterised. Through understanding how the duty cycle of the observations affected the power spectrum informed the way we finally parametrised the full model of the power spectrum.

5.4.2 Parametrisation of the SMMF Power Spectrum

In the frequency domain, each of the RM peaks models well as a Lorentzian distribution, similar to peak-bagging modes of solar oscillations (Handberg & Campante, 2011; Davies et al., 2014), which is due to the quasi-coherent nature of the source. The exponential decay of the RM SMMF source gives width to the peaks in the power spectrum, which we can measure to infer their lifetime.

A single, symmetric Lorentzian peak can be modelled by:

$$L_n(\nu; \Gamma, A_n, \nu_n) = \frac{2A_n^2}{\pi\Gamma} \left(1 + \left(\frac{\nu - \nu_n}{\Gamma/2} \right)^2 \right)^{-1}, \quad (5.10)$$

where ν is frequency, A_n is the Root Mean Square (RMS) amplitude of the RM component in the time-domain, Γ is the linewidth of the RM peak, ν_n is the frequency of the RM peak, and n simply flags each peak. The mean-squared power in the time

domain from the RM component of the SMMF is given by the sum of the A_n^2 of the individual harmonics in the power spectrum.

Upon closer inspection of the power spectrum it is possible to see that the peaks appear to exhibit an asymmetric shape (see Fig. 5.2 and Fig. 5.3). Taking inspiration from Howe et al. (2020), it is possible to allow for asymmetry, which is controlled by the asymmetry parameter, α (Stancik & Brauns, 2008):

$$L_n(\nu; \Gamma, A_n, \nu_n) = \frac{2A_n^2}{\pi\Gamma(\nu)} (1 + (2X(\nu))^2)^{-1}, \quad (5.11)$$

where:

$$X(\nu) = (\nu - \nu_n)/\Gamma(\nu); \quad (5.12)$$

$$\Gamma(\nu) = 2\Gamma/[1 + \exp^{-\alpha(\nu - \nu_n)}]. \quad (5.13)$$

In the limit where $\alpha \rightarrow 0$, we see that $\Gamma(\nu) \rightarrow \Gamma$, thus the asymmetric expression equates to the symmetric expression.

The model function used to describe the RM signal in the power spectrum is given by equation (5.14); the sum of N Lorentzian-peaks:

$$P(\nu) = \sum_{n=1}^N L_n(\nu; \Gamma, A_n, \nu_n). \quad (5.14)$$

The subscript, n , describes a single peak in the power spectrum; in implementing the model we constrain the central frequency for each of the peaks such that they must be integer values of ν_0 : $\nu_n = n\nu_0$. This means that we define a single rotation frequency only, and subsequent peaks are harmonics. It is worth noting explicitly that this function assumes the linewidth of each Lorentzian peak is the same; only their amplitudes and central frequencies differ.

When modelling the power spectrum we attempted the fit with both of the symmetric and asymmetric Lorentzian expressions, independently. Firstly we modelled using the symmetric Lorentzian, and subsequently, using the asymmetric Lorentzian.

This would determine whether there is a necessity for the addition of the asymmetry parameter.

Through both formulations we can measure the *e*-folding lifetime of the amplitude of the RM component (T_e), as it is related to the linewidth of the peak by:

$$\Gamma = (\pi T_e)^{-1}. \quad (5.15)$$

The low-frequency power due to instrumental drifts and solar activity can be reasonably well captured by the inclusion of a zero-frequency centred Lorentzian, i.e. Harvey-function, given by:

$$H(\nu; \sigma, \tau) = \frac{4\sigma^2\tau}{1 + (2\pi\nu\tau)^2}, \quad (5.16)$$

where σ is the characteristic amplitude of the low frequency signal, and τ describes the characteristic timescale of the excursions around zero in the time-domain.

Finally, the high frequency power is accounted for by the inclusion of a constant offset due to shot-noise, c .

5.4.3 Modelling the SMMF Power Spectrum

Parameter estimation was performed in a Bayesian manner using a Markov Chain Monte Carlo (MCMC) fitting routine. Following from Bayes theorem we can state that the posterior probability distribution, $p(\mathbf{a}|D, I)$, is proportional to the likelihood function, $L(D|\mathbf{a}, I)$, multiplied by a prior probability distribution, $p(\mathbf{a}|I)$:

$$p(\mathbf{a}|D, I) \propto L(D|\mathbf{a}, I) p(\mathbf{a}|I), \quad (5.17)$$

where D are the data, and I is any prior information.

To perform the MCMC integration over the parameter space we must define a likelihood function; however, in practice, it is more convenient to work with log-

arithmic probabilities. The noise in the power spectrum is distributed as χ^2 2 degrees-of-freedom (Handberg & Campante, 2011; Davies et al., 2014), therefore the log likelihood function is:

$$\ln(L) = - \sum_i \left\{ \ln(M_i(\mathbf{a})) + \frac{O_i}{M_i(\mathbf{a})} \right\}, \quad (5.18)$$

for a model, M_i , with parameters, \mathbf{a} , and observed power, O_i , where i describes the frequency bin. This likelihood function assumes that all the frequency bins are statistically independent but the effect of the window function means that they are not. We handled this issue by using simulations based on the artificial data discussed in Section 5.5.1.

The prior information on each of the parameters used during the MCMC is discussed below, in Section 5.5.2.

The affine-invariant MCMC sampler `emcee` (Foreman-Mackey et al., 2013) was employed to explore the posterior parameter space. The chains are not independent when using `emcee`, therefore convergence was interrogated using the integrated autocorrelation time. We computed the autocorrelation time and found $\tau \sim 120$ steps. Foreman-Mackey et al. (2013) suggests that chains of length $\geq 50\tau$ are often sufficient. After a burn in of 6000 steps, we used 7000 iterations on 50 chains to explore the posterior parameter space, which was sufficient to ensure we had convergence on the posterior probability distribution.

5.4.4 Comparison with the WSO SMMF

To provide comparative results on the inferences from the BiSON SMMF, we repeated the analysis on the power spectrum of the WSO SMMF. The WSO data are only provided on a daily cadence, hence the Nyquist frequency is lower than for 40-second BiSON data, at ~ 5.79 μ Hz, and it was not possible to observe the SB component. The analysis was therefore also repeated using the daily-averaged BiSON SMMF, to provide a more direct comparison between BiSON and WSO.

The same parametrisation as outlined above was relevant to the modelling of the features in the WSO Power Spectral Density (PSD), and the RM peaks were fitted using a model with symmetric Lorentzian peaks and separately with asymmetric Lorentzian peaks.

5.5 Results

5.5.1 Investigation of the Window Function

Due to the low fill of data, we see the effects of daily and random gaps on the power spectrum. Periodic gaps in the data give rise to sidebands in the power spectrum and random gaps cause a more broadband shifting of power, meaning that some power from the low-frequency RM component in the power spectrum is aliased to higher frequencies. First we will concentrate specifically on the effect of daily, periodic gaps in the data. The daily, periodic gaps in the BiSON data, due to single-site observations, produce sidebands around a frequency of $1/\text{day} \sim 11.57 \mu\text{Hz}$ and its harmonics.

The frequency (and harmonics) of the RM component are located near zero ($\nu_0 \sim 0.4 \mu\text{Hz}$). We are usually only interested in the real, positive frequencies but due to their close proximity to zero, they are reflected back as a product of the aliasing and hence there are negative and positive side-bands in the (positive frequency) power spectrum. When considering the aliased power, both the positive and negative side-bands must be taken into account. The aliased power is located at frequencies:

$$\nu_{n,i} = i \left(\frac{1}{\text{day}} \pm \nu_n \right), \quad (5.19)$$

where i denotes the side-band number, and n denotes the harmonic of the mode.

The side-band structure implied by equation (5.19) is shown clearly in the SMMF power spectrum in Figure 5.4. It is clear that we could therefore have used the predicted locations of the aliased power and incorporated them into the model for the

full power spectrum. This would, however, have required us to explicitly model some ~ 1100 groups of side-bands in order to cover this effect over the entire frequency range, and each group would have required a unique parameter to control the fraction of power that was contributed to the full PSD. It would have become computationally expensive to model each aliased peak and there would certainly have been room for degeneracy issues to occur.

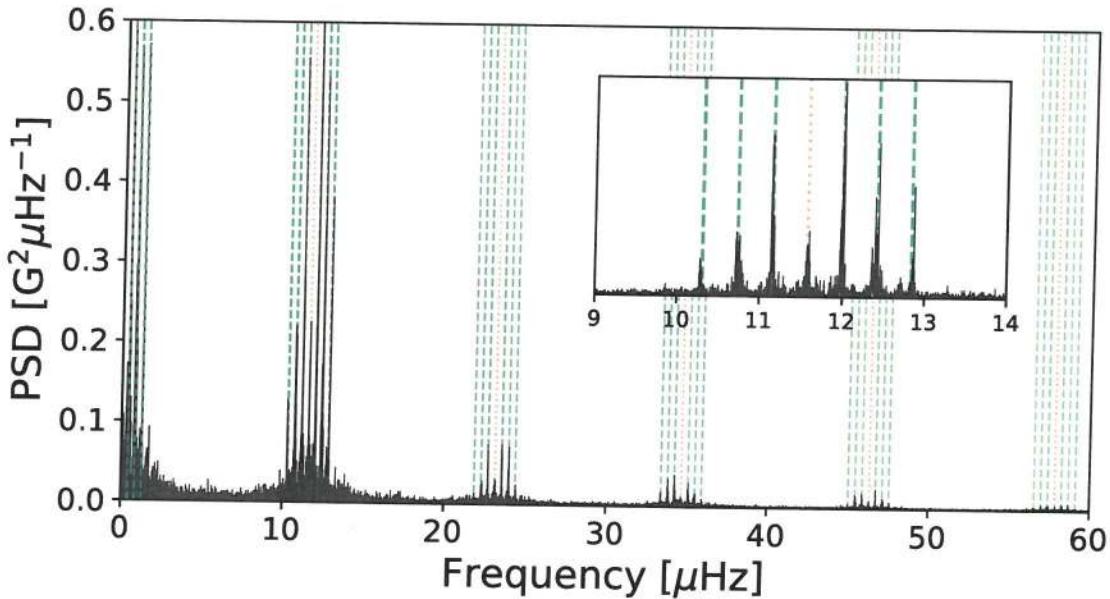


Figure 5.4: Locations of aliased power in side-band peaks. The orange, dotted-lines show the locations of frequencies at multiples of 1/day. The green, dashed-lines show the locations of the side-band peaks – harmonic frequencies reflected around multiples of 1/day. The inset shows a zoom of one set of side-band peaks around 1/day.

A more desirable, and we shall see more accurate, approach was to utilise the power spectrum of the window function itself. This approach not only takes into account the effect of daily, periodic gaps in the data, but also the more complex features that manifest in the power spectrum due to the random gaps in the data. To do this, the Fourier transform of the window function describing the duty cycle of observations was computed (i.e. $|\mathcal{F}[g(t)]|^2$), where the duty cycle function, $g(t)$, is given by:

$$g(t) = \begin{cases} 1 & \text{for } |B(t)| > 0 \\ 0 & \text{for } |B(t)| = 0 \end{cases}. \quad (5.20)$$

In Figure 5.5 the power spectrum of the window function is shown. Furthermore, to demonstrate the effect of the window function on the power spectrum, an artificial spectrum was simulated with a single Lorentzian peak, following equation (5.14). By computing the inverse Fourier transform, an artificial time-series was generated over the same epoch as the BiSON SMMF observations. We were then able to examine the effects of injecting gaps into the data that were concurrent with the BiSON SMMF gaps.

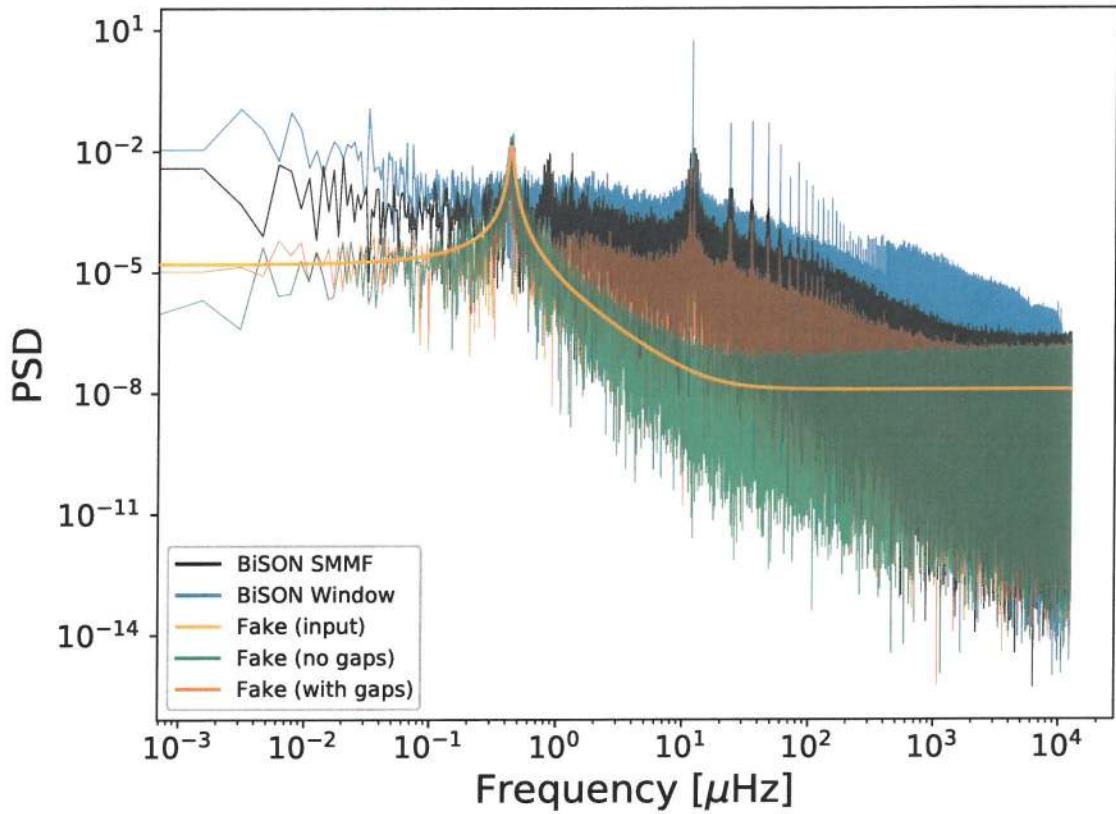


Figure 5.5: The effects of the window function on the power spectrum is shown by using a fake data set and this is compared to the BiSON power spectrum. Black line: BiSON SMMF PSD; blue line: power spectrum of the window function; green and dark-orange lines: the power spectrum of the artificial data without and with gaps, respectively; light orange line: the input peak used to generate the artificial data over-plotted. The power spectra of the BiSON SMMF and the window function have been shifted upwards by a factor of 6 and 30, respectively, for clarity.

As well as the power spectrum of the window function, Figure 5.5 also shows the underlying noise-free peak that describes the fake data and the power spectra of the artificial data with and without the injected gaps. The power spectrum of the

BiSON SMMF data is also plotted for comparison.

It is strikingly clear from Figure 5.5 that the shape of the spectrum of the window function has a remarkable resemblance to the BiSON SMMF spectrum and the output of the artificial spectrum with gaps injected. This demonstrates that the periodic window function, with such a low duty cycle, has a significant effect on the power spectrum of the input signal, which not only produces the diurnal sidebands, but also a broadband spread of the power.

Due to the broadband shape of the window function power spectrum compared to the BiSON SMMF, it appears that there is actually no detectable red-noise-like, SB component in the SMMF; it is instead a manifestation of the periodic gaps in the data.

To analytically understand this effect, we can express the time series of observed data ($y(t)$) as a multiplication of the uninterrupted, underlying signal ($f(t)$) with the window function ($g(t)$), i.e.:

$$y(t) = f(t) g(t). \quad (5.21)$$

In the frequency domain, the Fourier transform of a product becomes the convolution of the transformed components. It is possible to express the observed power spectrum of data with periodic gaps in terms of the window function and the gap-free power spectrum, given by:

$$P'(\nu; \mathbf{a}) = P(\nu; \mathbf{a}) * |\mathcal{F}[g(t)]|^2. \quad (5.22)$$

Therefore to model the observed power spectrum in a robust manner, which takes into account the intricacies caused by gaps in the data, we used a model which was formed of a model power spectrum, $P(\nu; \mathbf{a})$, convolved with the Fourier transform of the window function describing the duty cycle of observations ($|\mathcal{F}[g(t)]|^2$), i.e. a

model described by equation (5.22), where:

$$P(\nu; \mathbf{a}) = \sum_{n=1}^N L_n(\nu; \Gamma, A_n, \nu_n) + H(\nu; \sigma, \tau) + c. \quad (5.23)$$

Care was taken to ensure Parseval's theorem was obeyed, and no power was lost or gained from the convolution operation:

$$\sum_{\nu} P'(\nu) = \sum_{\nu} P(\nu) = \frac{1}{N} \sum_t B(t)^2, \quad (5.24)$$

where N here is the number of observed cadences.

To demonstrate this method, as a proof-of-concept, we fit a model of a single Lorentzian peak, plus a shot-noise background, to the gap-free fake PSD (without the convolution) and the fake PSD using the gaps (requiring the convolution). The modelling was performed using the affine-invariant MCMC sampler `emcee` (Foreman-Mackey et al., 2013) to explore the posterior parameter space, and the results of this fit are summarised in Table 5.1.

Table 5.1: Model parameter values for the generation of artificial data, and the median posterior values for the fit to the power spectra generated with and without the gaps in the data. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.

Parameter	Input	Fit (no gaps)	Fit (gaps)	Unit
ν_0	0.42867	$0.4277^{(+18)}_{(-18)}$	$0.4261^{(+03)}_{(-03)}$	μHz
Γ	0.030	$0.0279^{(+37)}_{(-36)}$	$0.0340^{(+06)}_{(-06)}$	μHz
A	100.0	$101.2^{+7.3}_{-6.0}$	111.7 ± 0.1	mG
c	0.20	$0.1872^{(+03)}_{(-03)}$	$0.1876^{(+01)}_{(-01)}$	G^2Hz^{-1}

As we can see from the results in Table 5.1, the median values of the parameter posterior distributions without the gaps are in accordance with the input values, which were used when generating the artificial data, within uncertainties. When the gaps were introduced, the median values of the parameter posterior distributions were close, but the widths of the distributions (which give implied confidence

intervals) did not overlap.

Through this demonstration we observed that, as a result of the convolution in the model, the widths of the posterior distributions for the model parameters were systematically underestimated. This effect arises because we do not account explicitly for the impact of the window function convolution on the covariance of the data. To perform the full likelihood evaluation, with the correlated noise, requires large-data computational linear algebra (i.e. the inversion of an N-by-N diagonal-constant/Toeplitz matrix, where N is the size of data); unfortunately the process of fully accounting for the correlated noise in this scenario was too computationally expensive, due to the large data set with $\sim 10^7$ data points. Our ability to resolve the model parameters is well represented in the case where the convolution was not performed; this helped us to understand how the convolution affected our ability to measure the true posterior widths, which allowed us to account for the systematic underestimate of the credible regions of the posterior distributions. We assumed the for ν_0 and c , we have an accuracy of $\sim 0.5\%$ each; for Γ , an accuracy of $\sim 15\%$; for A , an accuracy of $\sim 10\%$. Using these uncertainty estimates, the parameter posterior distributions for the convolution model are in accordance with the input values.

When modelling the power spectrum of the observed BiSON SMMF, these factors allowed us to account for the systematic underestimate of the posterior widths from the convolution.

5.5.2 Modelling the BiSON Power Spectrum

As there were many data points in the power spectrum, each likelihood calculation was computationally expensive. In order to reduce the required computation, the BiSON power spectrum was cut at a frequency of 7000 μ Hz, as at very high frequency, the spectrum purely represents the shot noise in the SMMF, and it was deemed as a sufficient limit to still fully converge on the shot noise parameter.

The BiSON power spectrum was modelled against equation (5.22) (which used

equation (5.23) with $N = 4$ peaks), for both the symmetric and asymmetric Lorentzian models, using the affine-invariant MCMC sampler `emcee` (Foreman-Mackey et al., 2013) to explore the posterior parameter space. In the modelling we used uniform prior information, providing reasonable boundaries on each parameter, as detailed below.

$$\nu_0 \sim \mathcal{U}(0.38, 0.50) \text{ } \mu\text{Hz}$$

$$\Gamma \sim \mathcal{U}(0.00, 0.11) \text{ } \mu\text{Hz}$$

$$A_1 \sim \mathcal{U}(100, 350) \text{ mG}$$

$$A_2 \sim \mathcal{U}(50, 200) \text{ mG}$$

$$A_3 \sim \mathcal{U}(20, 150) \text{ mG}$$

$$A_4 \sim \mathcal{U}(10, 100) \text{ mG}$$

$$\sigma \sim \mathcal{U}(0.01, 500) \text{ mG}$$

$$\tau \sim \mathcal{U}(0.10, 200) \text{ } 10^6 \text{ s}$$

$$c \sim \mathcal{U}(10^{-3}, 10^2) \text{ G}^2 \text{ Hz}^{-1}$$

$$\alpha \sim \mathcal{U}(-500, 0)$$

In Table 5.2 the median values of adjusted, marginalised posterior distributions for each of the model parameters are displayed, for both the symmetric and asymmetric models. The resultant posterior distributions were approximately normally distributed and there was no significant covariance between parameters, therefore uncertainties on the parameters correspond to the 68% credible intervals either side of the median. We note the we have previously shown the convolution results in a systematic underestimate in the width of the posterior distributions and thus here we are presenting an estimation of the true uncertainties based on multiplying by

the factors acquired from the artificial simulations in Section 5.5.1.

Table 5.2: Median values of the marginalised posterior distributions for each model parameter in the fit to the BiSON power spectrum using symmetric and asymmetric Lorentzian profiles. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median in the adjusted posteriors. The last row in the table shows the Bayesian Information Criterion (BIC) value for each model.

Parameter	40-s symm.	40-s asymm.	Unit
ν_0	$0.4270^{(+18)}_{(-18)}$	$0.4278^{(+18)}_{(-18)}$	μHz
Γ	$0.0264^{(+35)}_{(-35)}$	$0.0316^{(+41)}_{(-41)}$	μHz
A_1	166.0 ± 10.7	178.9 ± 11.5	mG
A_2	115.9 ± 7.4	129.0 ± 8.3	mG
A_3	83.2 ± 5.3	93.5 ± 6.0	mG
A_4	32.6 ± 2.1	38.9 ± 2.5	mG
τ	51.8 ± 6.8	62.7 ± 8.2	days
σ	83.4 ± 5.4	79.1 ± 5.1	mG
c	$0.2103^{(+03)}_{(-03)}$	$0.2102^{(+03)}_{(-03)}$	$\text{G}^2 \text{Hz}^{-1}$
α	—	-119.8 ± 15.7	—
BIC	106	122	—

We calculated the Bayesian Informatin Criterion (BIC) to aid our model selection, to determine whether the asymmetry was truly necessary or a case of overfitting to the data. The BIC for both the symmetric and asymmetric models were ~ 106 and ~ 122 , respectively. This highlighted that the model using the symmetric Lorentzian profiles was favoured, due to the lower BIC value.

The convolved model of the data, using symmetric Lorentzian peaks, is shown in Figure 5.6, over-plotted on top of the BiSON SMMF power spectrum.

The central frequency of this model, ν_0 , implies a synodic rotation period of 27.11 ± 0.11 days, and hence a sidereal rotation period of 25.23 ± 0.11 days. The rotation period measured is in agreement with other literature values for the rotation signal in the SMMF (Chaplin et al., 2003; Xie et al., 2017), and is in accordance with that typically observed for ARs and sunspots.

According to the model for differential rotation given by Snodgrass (1983) and Brown et al. (1989), the measured rotation period implies the RM component of

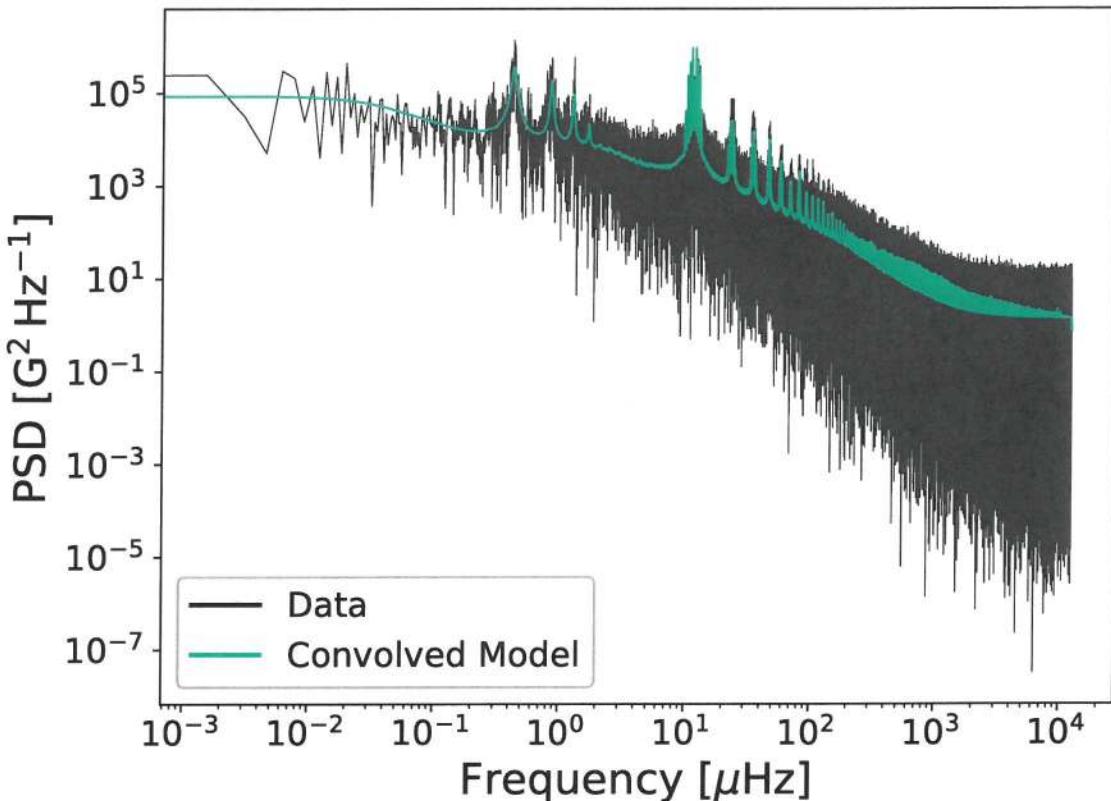


Figure 5.6: Full, modelled power spectrum of the BiSON SMMF on logarithmic axes. The data are displayed in black and the convolved model using symmetric Lorentzian peaks is shown in green.

the SMMF is sensitive to a time-averaged latitude at around 12° . This latitude is consistent with those spanned by sunspots and ARs over the solar activity cycle (Maunder, 1904; McIntosh et al., 2014), and particularly during the declining phase of the solar cycle (Thomas et al., 2019). This strongly implies that the origin of the RM component of the SMMF is linked to ARs and MFCs.

Furthermore, from the measured linewidth of the Lorentzian peaks, we have calculated the lifetime of the RM component using equation (5.15). The linewidth suggests a lifetime of 139.6 ± 18.5 days, which is in the region of $\sim 20 \pm 3$ weeks. The typical lifetime of ARs and sunspots is usually on the order of weeks to months, dependent on their size (Zwaan, 1981; Schrijver & Harvey, 1994; Howard, 2001; Hathaway & Choudhary, 2008; van Driel-Gesztelyi & Green, 2015), therefore we have measured a lifetime of the RM component which is consistent with the lifetime

of ARs and sunspots. This again suggests that the source of the signal is linked to active regions of magnetic field or similar MFCs.

Taking into account the work performed by [Bose & Nagaraju \(2018\)](#), which showed evidence to suggest sunspots did not contribute to the SMMF, and also our concerns with their methodology, we are cautious to specify that sunspots are the source of the RM component; however, this can still not be ruled out altogether. The method of identifying ARs or strong MFCs in magnetograms by [Bose & Nagaraju \(2018\)](#) potentially mis-identified regions of magnetic flux associated with ARs and MFCs as background flux and it is possible that they do contribute to the SMMF. By comparison with the work carried out by [Kutsenko et al. \(2017\)](#), the work performed in this project agrees that the SMMF is dominated by features with properties in-line with spots, ARs, and MFCs, i.e. long-lived and making up a small fraction of the solar disc, confined to active bands of latitude.

With all this considered, we conclude that our investigation of the BiSON SMMF indicates the SMMF has its origin in the vicinity of ARs and other concentrations of strong flux that are long-lived on the solar disc and exist in active latitudes. Whether or not specifically this is due to spots or other MFCs will require further work on the magnetogram thresholding techniques investigating the SMMF.

5.5.3 Comparison to the WSO Power Spectrum

Here we present the results from modelling the WSO power spectrum; we also show the results of modelling the power spectrum of the daily-averaged BiSON spectrum, to provide a direct comparison between both observations on the same cadence.

The fill of the WSO observations over the same epoch as BiSON observations was $\sim 78\%$, and at $\sim 55\%$ for the daily-averaged BiSON observations. As the fill of the daily-averaged BiSON data was higher than the 40-cadence data, the effect of the window function was less prominent. Nevertheless, the power spectra for both daily WSO and BiSON data were modelled using the convolution approach, as per the

previous section. As, compared to the duty cycle of 40-cadence data, the fill is closer to 100%, we also expect the effect of the convolution to be less significant on the widths of the posterior distributions for each parameter, and thus the uncertainties to be more representative.

In addition, we also had to account for the sinc-function attenuation effect and reflections of power around the Nyquist frequency, due to the lower Nyquist frequency and its close proximity to the frequencies of the rotation peaks (Basu & Chaplin, 2017).

The daily-averaged BiSON power spectrum was modelled against equation (5.22) using equation (5.23) with $N = 4$ peaks, and similarly for the WSO power spectrum but with $N = 3$ peaks, as a fourth peak was less pronounced. Both fitting routines used the affine-invariant MCMC sampler `emcee` (Foreman-Mackey et al., 2013) to explore the posterior parameter space, using 10000 iterations on 50 chains. The convergence was again interrogated using the integrated autocorrelation time, to ensure a sufficient number of effective samples were used; hence ensuring that the posterior distribution was sampled.

In Table 5.3 the median values of marginalised posterior distributions for the model parameters are displayed, for both daily WSO and BiSON data. Reported uncertainties on the parameters correspond to the 68% credible intervals either side of the median. The systematic underestimate of the posterior width is much less of a concern with these results due to the higher duty cycle; the uncertainties are on the same order as the results not using a convolved model in Table 5.1. The convolved model for both the WSO and BiSON data, using symmetric Lorentzian peaks, are shown in Figure 5.7 over-plotted on top their respective SMMF power spectra.

The results of the fit to the daily-averaged BiSON data are similar to those for the 40-second data, however there are a few differences which arise due to the different window functions and the realisations of the noise. We do however, generally, see

Table 5.3: Median values of the marginalised posterior distributions for each model parameter in the fit to the daily WSO and BiSON power spectra. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median. The last row in the table shows the BIC value for each model.

Parameter	WSO		BiSON 24-hr		Unit
	Symm.	Asymm.	Symm.	Asymm.	
ν_0	$0.4272^{(+05)}_{(-05)}$	$0.4290^{(+04)}_{(-04)}$	$0.4246^{(+08)}_{(-09)}$	$0.4270^{(+08)}_{(-08)}$	μHz
Γ	$0.0210^{(+16)}_{(-15)}$	$0.0210^{(+16)}_{(-15)}$	$0.0359^{(+27)}_{(-26)}$	$0.0365^{(+30)}_{(-29)}$	μHz
A_1	$134.8^{+3.1}_{-3.2}$	137.3 ± 3.5	$118.9^{+4.2}_{-3.8}$	$125.6^{+5.0}_{-4.4}$	mG
A_2	121.6 ± 3.4	$125.9^{+3.7}_{-3.6}$	$109.2^{+3.6}_{-3.8}$	$115.4^{+4.2}_{-4.4}$	mG
A_3	$75.1^{+3.5}_{-3.4}$	$78.9^{+4.0}_{-3.9}$	$64.6^{+3.7}_{-3.5}$	$70.1^{+4.2}_{-4.0}$	mG
A_4	—	—	$28.0^{+3.4}_{-3.3}$	$30.7^{+3.7}_{-3.6}$	mG
τ	$263.9^{+121.5}_{-83.3}$	$270.8^{+128.5}_{-84.5}$	$49.8^{+18.5}_{-12.7}$	$77.5^{+32.4}_{-22.0}$	days
σ	52.3 ± 4.6	$51.3^{+4.6}_{-4.5}$	$81.5^{+5.1}_{-1.1}$	$77.1^{+5.7}_{-5.5}$	mG
c	$11.7^{+18.7}_{-8.8}$	$11.0^{+18.4}_{-8.4}$	0.210 ± 0.011	0.210 ± 0.011	$\text{G}^2 \text{ Hz}^{-1}$
α	—	$-61.7^{+13.3}_{-15.3}$	—	$-42.1^{+12.0}_{-15.1}$	—
BIC	47	55	55	63	—

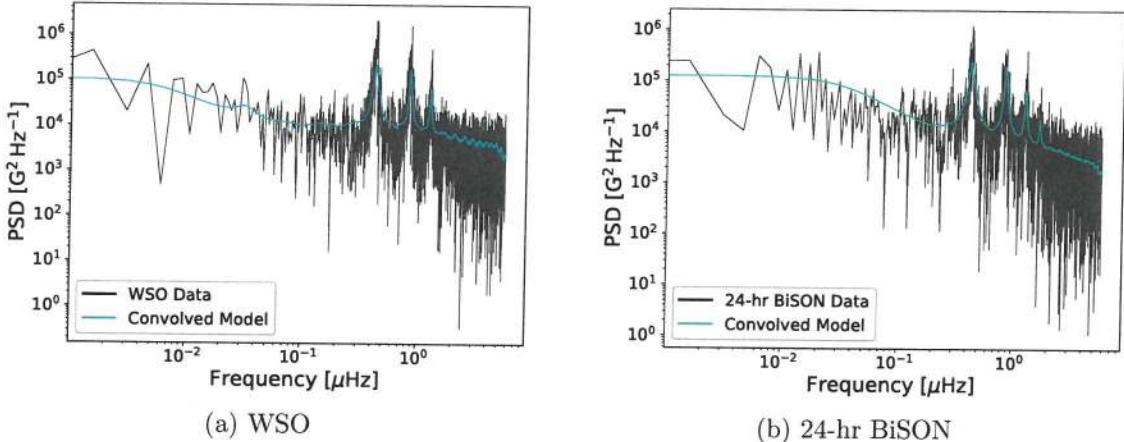


Figure 5.7: Modelled power spectrum of (a) the WSO SMMF; (b) the daily-averaged BiSON SMMF, on logarithmic axes. The data are displayed in black and the convolved model using asymmetric Lorentzian peaks is shown in blue and green, for WSO and BiSON, respectively.

a good agreement between the parameters. We see in Figure 5.8 that there are differences between the daily-averaged and 40-second spectra; at low frequencies it is possible to see the differences in the realisations of the noise, and at higher frequencies we can see differences due to the window function aliasing. This plot shows the possible reasons why the parameters in Table 5.3 may slightly differ from

the results presented in Table 5.2.

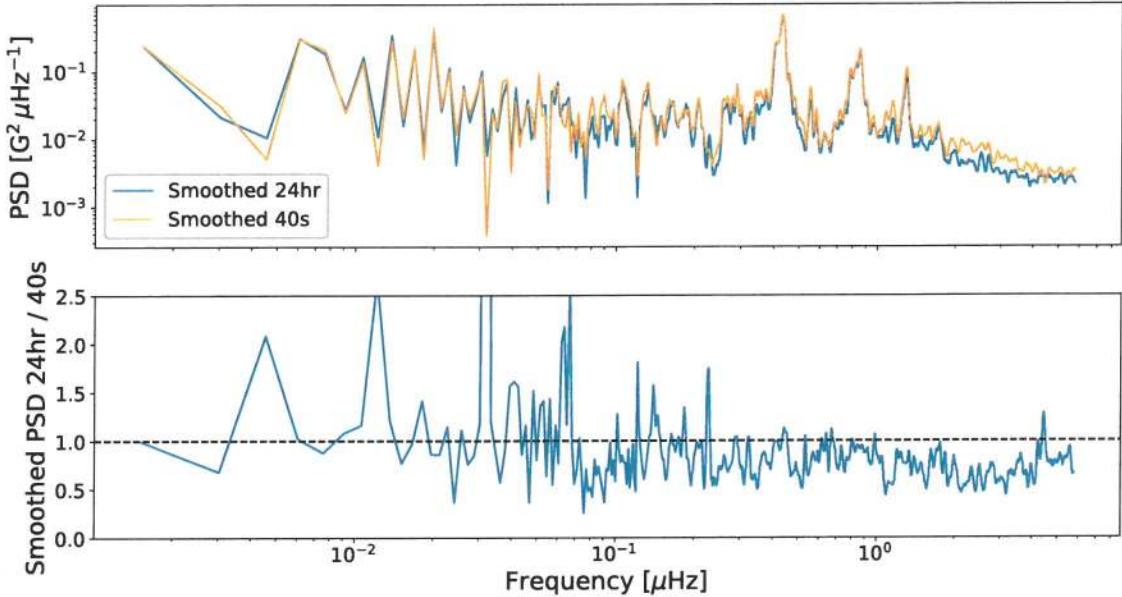


Figure 5.8: A comparison between the power spectra produced using the daily-averaged BiSON data and the 40-second cadence BiSON observations. The top plot shows the log-smoothed power spectra of the daily-averaged data (blue) and the 40-second data (orange). The bottom panel show the ratio of the daily-averaged data power spectrum to the 40-second data power spectrum. The horizontal, dashed line indicates a ratio of 1.

The fit to the WSO power spectrum using a model with asymmetric Lorentzian profiles converged on an asymmetry parameter, but as with the 40-second analysis of the BiSON data, we calculated the BIC values for both the symmetric and asymmetric model to determine which model to select. The BIC values for the WSO models using symmetric and asymmetric Lorentzian profiles were ~ 47 and ~ 55 , respectively, and the BIC values for the BiSON models using symmetric and asymmetric Lorentzian profiles were ~ 55 and ~ 63 , respectively. In both cases, this highlighted that the models using the symmetric Lorentzian profiles were favoured, due to the lower BIC values.

The rotation period in the WSO data is in agreement with that measured using BiSON data to within 3σ , and this period implies a cycle-averaged latitude of around $\sim 12^\circ$. This agrees with the conclusions drawn from our inferences of the BiSON data, that the RM source is linked to ARs.

The linewidth from the WSO data implies a RM lifetime of 175 ± 13 days, which is in the region of ~ 25 weeks, or half a year. The WSO lifetime is in agreement with the 40-second BiSON data. However, this lifetime is inconsistent with that measured using the daily-averaged BiSON data. The lifetime measured using the daily-averaged BiSON data is around 50% smaller than that measured with the WSO data. A possible reason for this could be the differences arising from systematics in the fitting. When using the daily-averaged BiSON data it is harder to accurately model the background because of the lower Nyquist frequency, hence we may be vulnerable to systematics impacting on the inferred widths as a higher fitted background would tend to produce narrower peak widths. These limits are still consistent however with the lifetime of large, strong ARs (Schrijver & Harvey, 1994; van Driel-Gesztelyi & Green, 2015).

As with the BiSON observations, the investigation of the WSO SMMF also indicates the origin of the SMMF is linked to ARs and MFCs that are long-lived on the solar disc and exist in active latitudes.

5.6 Discussion

5.6.1 Testing the Effects of Differential Rotation and Active Region Migration

We know the rotation period of ARs varies throughout the solar cycle as a result of solar differential rotation and latitudinal migration. As we have inferred that the RM component of the SMMF is likely linked to ARs and MFCs, we may therefore assume that the RM component is also sensitive to these effects. Here we analyse the effect of migration and differential rotation on our ability to make inferences on the lifetime of the RM component.

Several studies have modelled the solar differential rotation, and its variation with latitude and radius of the Sun (see Beck, 2000; Howe, 2009, for an in depth

review of the literature on solar differential rotation). Magnetic features have been shown to be sensitive to rotation deeper than the photosphere; therefore, in general, magnetic features can be seen to rotate with a shorter period than the surface plasma (Howe, 2009).

Chaplin et al. (2008) analysed the effects of differential rotation on the shape of asteroseismic low- l p modes of oscillation, and showed that the consequence of differential rotation is to broaden the observed linewidth of a mode peak. The authors provide a model of the resultant profile of a p mode whose frequency is shifted in time to be a time-average of several instantaneous Lorentzian profiles with central frequency $\nu(t)$, given by:

$$\langle P(\nu) \rangle = \frac{1}{T} \int_0^T H \left(1 + \left(\frac{\nu - \nu(t)}{\Gamma/2} \right)^2 \right)^{-1} dt. \quad (5.25)$$

The angled brackets indicate an average over time. H and Γ are the mode height (maximum power spectral density) and linewidth, respectively. The full period of observation is given by T .

Chaplin et al. (2008) also show that by assuming a simple, linear variation of the unperturbed frequency, ν_0 , from the start to the end of the time-series by a total frequency shift $\Delta\nu$ (see equation (5.26)),

$$\nu(t) = \nu_0 + \Delta\nu \frac{t}{T}, \quad (5.26)$$

the resultant profile of a p mode can analytically be modelled by:

$$\langle P(\nu) \rangle = \frac{H}{2\epsilon} \arctan \left(\frac{2\epsilon}{1 - \epsilon^2 + X^2} \right), \quad (5.27)$$

where ϵ and X are defined in equation (5.28) and equation (5.29):

$$\epsilon = \frac{\Delta\nu}{\Gamma}; \quad (5.28)$$

$$X = \frac{\nu - [\nu_0 + (\Delta\nu/2)]}{\Gamma/2}. \quad (5.29)$$

As the mode linewidths are broadened by this effect, we evaluated whether our ability to resolve the true linewidth of the RM component, and hence its lifetime, was affected. To evaluate this we computed the broadened profiles given by both equation (5.25) and equation (5.27), and fit the model for a single Lorentzian peak, to determine whether the linewidth is recovered.

In the first instance, we computed the broadened peak using equation (5.25). Over the duration of the observations, we computed the daily instantaneous profile, $P(\nu(t))$. The time-averaged profile, $\langle P(\nu) \rangle$, is a weighted average of each instantaneous profile, where the weights are given by the squared, daily-averaged SMMF, in order to allow a larger broadening contribution at times when the SMMF amplitude is large.

In the second instance, we computed the broadened peak using equation (5.27). Over the duration of the observations the daily frequency shift is computed, $\Delta\nu$. The time-averaged shift, $\Delta\nu$, is a weighted average, where again the weightings are given by the squared, daily-averaged SMMF.

To determine the shift in the rotation rate with migration, we used the model of the solar differential rotation as traced by magnetic features (Ω_m) given by equation (5.30), where $\mu = \cos\theta$ and θ is the co-latitude (Snodgrass, 1983; Brown et al., 1989):

$$\frac{\Omega_m}{2\pi} = 462 - 74\mu^2 - 53\mu^4 \text{ nHz}. \quad (5.30)$$

Finally, the time-dependence on the latitude of the active regions used the best-fitting quadratic model by Li et al. (2001).

In both instances, the broadened peak was modelled as a single Lorentzian peak using equation (5.10), with a width equivalent to that which was inferred from modelling the BiSON power spectrum. We used `emcee` (Foreman-Mackey et al., 2013) to explore the posterior parameter space with priors similar to the fit to the

full power spectrum.

Over the entire duration of the SMMF observations, the time-averaged profile was calculated, using equation (5.25), and this is shown in Figure 5.9a. The broadened mode used the input parameters for the model using symmetric Lorentzians, outlined in Table 5.2, however, with the background parameter set to zero.

By eye, the broadened profile does not appear to have a significantly larger linewidth. The input linewidth was $0.0264 \pm 0.0035 \mu\text{Hz}$, and the fit to the time-averaged broadened peak produced a linewidth of $0.0262^{+0.0038}_{-0.0037} \mu\text{Hz}$. The linewidth of the broadened peak under this method was rather unchanged from that of the true peak, and both linewidths are within uncertainties of each other.

Table 5.4: Input linewidth and the median posterior values of the Lorentzian model each simulation. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.

Input Value	Weighted Fit	Analytic Fit	Unit
0.0264 ± 0.0035	$0.0262^{+0.0038}_{-0.0037}$	$0.0263^{+0.0038}_{-0.0037}$	μHz

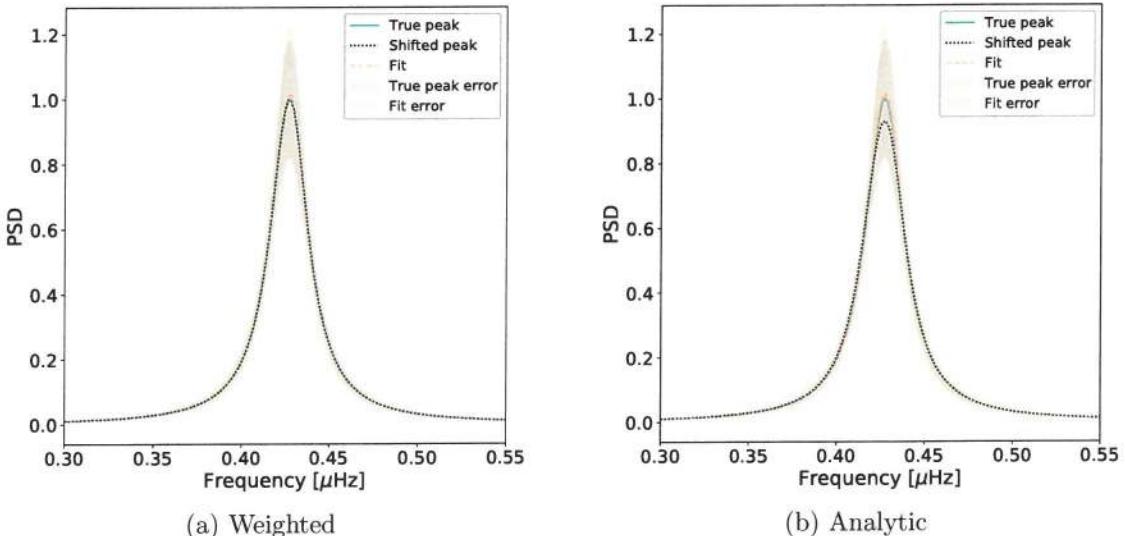


Figure 5.9: (a) Shows the Lorentzian distribution peak before and after the time-averaged broadening, and the fit to the broadened peak. (b) Shows the peak distribution before and after the analytical broadening, and the fit to the broadened peak. In both plots the broadened peaks have been shifted by the relevant frequency to overlay them on top of the true ν_0 for comparison.

The time-averaged frequency shift due to differential rotation was calculated, much in the same way as equation (5.25), to be $\Delta\nu = 0.01285 \mu\text{Hz}$. This shift was used to generate the broadened profile using equation (5.27). The broadened mode distribution also used the input parameters outlined in Table 5.2, however, with the background parameter set to zero.

Similar to the numerically broadened peak, by eye, the analytically broadened profile does not appear to have a significantly larger linewidth (see Fig. 5.9b). The input linewidth was $0.0264 \pm 0.0035 \mu\text{Hz}$, and the linewidth of the analytically broadened peak from the fit was $0.0263_{-0.0037}^{+0.0038} \mu\text{Hz}$, which was within the uncertainties of the linewidth of the input peak.

These results show that both numerically and analytically, the mode broadening effect of differential rotation and latitudinal migration does not affect our ability to resolve the linewidth of the peaks. Both broadening methods applied have been shown to have a negligible effect on the measured linewidth. This result provides confidence that the linewidth in Table 5.2 is the true linewidth of the RM peaks, thus providing the correct lifetime for RM component, unaffected by migration and differential rotation.

5.6.2 Further Morphology of the SMMF using SDO/HMI Data

In Chapter 6 we acquired SDO/HMI full-disc magnetograms, using the `SunPy` python module (Barnes et al., 2020), to support our investigations into Rossby waves (see next chapter for details).

Owing to having the SDO/HMI magnetograms, which provided the capability to separately analyse the Northern and Southern Hemispheres' MMF contribution to the SMMF during the rising phase of Cycle 24 in 2011 and during solar maximum in 2014, we also investigated whether there were hemispheric differences in the data, which resulted from the opposite polarities at high latitudes and towards

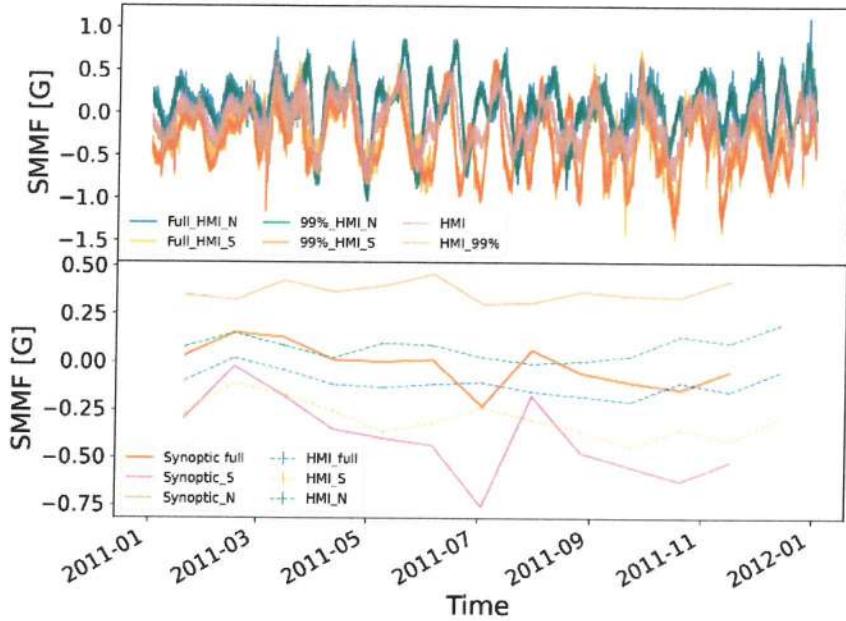
the poles. This served as a further analysis into other timescales which may exist in the SMMF. In particular, we investigated whether the SMMF exhibited an anti-correlation between the two hemispheres due to the oppositely polarised field near the polar regions, as found in synoptic charts, on a time-scale of the solar cycle.

To support this investigation, we acquired the synoptic charts from SDO/HMI. It was possible to average the signal over the Northern and Southern Hemispheres of the synoptic charts, as well as the full solar surface, thus providing a comparison to the hemispheric MMF and the full disc SMMF.

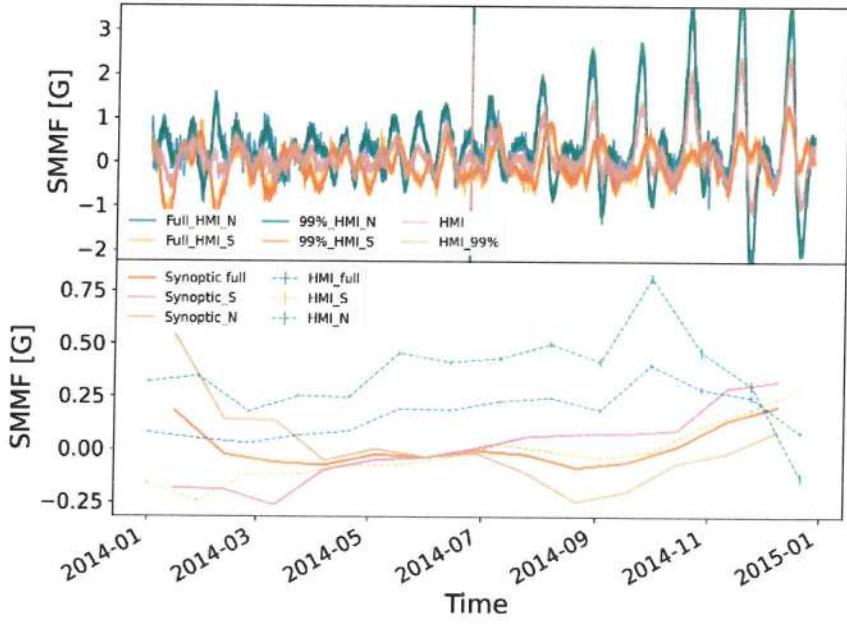
To compare the magnetogram data to the synoptic charts, we smoothed the separately averaged Northern and Southern Hemispheres' MMF and the full-disc SMMF signals using a box-car filter with a window width of a Carrington period, i.e. ~ 27 days. The resultant time series was plotted along with the hemispheric mean of the synoptic charts from SDO/HMI. Figure 5.10 shows the resultant smoothed hemispheric MMF and full-disc SMMF along with the average of the synoptic charts.

We can see from Figure 5.10 that there does exist a longer timescale in the hemispheric MMF when we average out the effects of the RM component. This is visible from the reversal of the field polarity in 2014. This longer timescale component resembles the average of the synoptic charts, and shows the secular variation which is contributed by the solar dipole at high latitudes. This timescale is the solar activity period and can be seen particularly in Figure 5.10b. We see the beginning of a the magnetic field reversal at around solar maximum, in mid-2014, with the onset of the “rush to the poles” after solar maximum (Wilson, 1994; McIntosh et al., 2019).

Interestingly, the field reversal was located during different epochs when comparing the synoptic chart data to the visible disc, hemispheric MMF data, and was delayed in the hemispheric MMF by around 7 Carrington rotations. Naturally, when the full-disc averaged SMMF was smoothed, the RM component was averaged out, resulting in a near-flat line. This was however expected as it was the average of the opposite hemispheres, and is consistent with the synoptic charts.



(a) 2011



(b) 2014

Figure 5.10: Investigations of timescales in the SDO/HMI magnetograms over 2011 and 2014. Both plots show in the top panel, the hemispheric MMF and full-disc SMMF from the magnetograms. The lower panel of each plot displays a comparison between the hemispheric and full-disc mean of the synoptic charts, compared to the box-car smoothed MMF from the magnetograms. N: Northern hemisphere; S: Southern hemisphere. Full HMI: considers the full solar disc; 99 HMI: considers only the inner 99% of the solar disc, by radius.

5.7 Conclusion

We have presented, for the first time, a frequency-domain analysis of over 20 years of high-cadence (40-second) BiSON observations of the SMMF.

Observations of the SMMF were computed from the Zeeman split D1 line of Potassium at ~ 770 nm, as measured by the Sutherland node of BiSON. The observations covered 7643 days over the period from 1992–2012 with a cadence of 40 seconds. A frequency-domain analysis of the SMMF was performed; the short cadence and long baseline of observations gave a fine frequency resolution in the power spectrum up to a high Nyquist frequency, allowing us to probe the elements that underpin the observed SMMF.

The duty cycle for the 40-second cadence observations was very low, hence the effect of the low fill on the power spectrum of the SMMF was investigated to help inform how to best model the full power spectrum. We highlighted that although there appeared to exist a red-noise-like, stochastic background component in the power spectrum, this was a feature originating from power aliasing due to the low duty cycle of the observations.

In the power spectrum, there was a strong peak at a frequency corresponding to the solar rotation, denoted the RM signal/component. It was also demonstrated that the low duty cycle aliased the power of the prominent peak due to the solar rotation to higher frequencies, which provided several copies of this peak at higher frequencies.

Using a model comprising of a series of Lorentzian peaks to model the RM signal, a Harvey function to account for lower frequency drifts, and shot-noise limit, which was convolved with the Fourier transform of the window function to account for the low duty cycle artefacts, we modelled the full power spectrum and measured the properties of the RM signal.

It was demonstrated that the convolution process affected the total power in the

model, thus careful treatment was taken to ensure Parseval's theorem was obeyed. In addition, it was shown that the width of the posterior distributions for the parameters had been systematically underestimated, as a result of the convolution process, because we do not account explicitly for the impact of the window function convolution on the covariance of the data. We could not resolve this, but simulated data provided a comparison between a model with and without convolution which was used to provide a correction to account for the systematic underestimate of the credible regions of the posterior when modelling the power spectrum of the observed BiSON SMMF.

A comparative study was conducted on the WSO data over the same observational epoch and the modelled power spectrum provide results that were in agreement with those measured in the BiSON power spectrum.

To further investigate the SMMF and our ability to infer the properties of the source, we used simulations to analyse the effects of differential rotation and AR migrations on our ability to measure the linewidth.

Finally, a short investigation into the SMMF as measured by SDO/HMI was conducted. Smoothing the data over the solar rotation period for both Northern and Southern hemispheres, separately, uncovered that the hemispheres display a longer variation, in accordance with the solar activity cycle, similar to that of the full-disc synoptic maps.

We leave the reader with the following points:

1. We have shown that there does not exist a short time-scale component in the SMMF, and the emergence of a red-noise-like signal in the power spectrum was due to the low duty cycle of the BiSON observations.
2. By modelling the peak of the RM signal as a symmetric Lorentzian profile, we found that the peak has a central frequency of $0.4270 \pm 0.0018 \mu\text{Hz}$. This measurement of the central frequency allowed us to infer the sidereal period of the RM signal to be 25.23 ± 0.11 days. This rotation suggests a magnetic

feature, cycle-averaged latitude of $\sim 12^\circ$, thus linking the source to active bands of latitude on the Sun.

3. The lifetime of the source of the RM component was inferred from the linewidth of the Lorentzian peaks to be 139.6 ± 18.5 days, which is in the region of $\sim 20 \pm 3$ weeks.
4. As a comparison, the power spectrum of the SMMF measured by WSO was also modelled and the linewidth and central frequency of the RM component were measured. The results were generally consistent with those from the BiSON data, and the conclusions inferred were in accordance.
5. The measured properties of the RM component of the SMMF are consistent with ARs. The literature advises that sunspots are not the origin of the SMMF, here we suggest that ARs and MFCs are the source of the dominant, rotation signal in the SMMF, that are long-lived on the solar disc and exist in active latitudes.
6. We have shown that our ability to determine the linewidth and hence lifetime of the RM modes was unaffected by AR migration and differential rotation.
7. Finally, a short investigation into the hemispheric contributions to the SMMF, using data from SDO/HMI, showed there is a longer-term variation which underpins the SMMF, in accordance with the activity cycle and polar field reversals.

At the time of writing, only two more of the BiSON nodes were actively measuring the SMMF (Las Campanas and Narrabri), and their measurements of the SMMF are not as stable as those measured by the Sutherland node. Sutherland has not been measuring the SMMF since 2013, however. Plans are in place to re-acquire these data in Sutherland and elsewhere, such that the frequency resolution can be

further increased with a longer baseline, allowing for more accurate inferences on the SMMF morphology.

With more time on the project, it would also be useful to develop a technique similar to Kutsenko et al. (2017) and Bose & Nagaraju (2018), which allows the SMMF to be dissected into regions and features on the disc.

6 Rossby Modes in the Solar Mean Magnetic Field

6.1 Introduction

Rossby waves, as first derived by Rossby & Collaborators (1939), have recently been discovered in the Sun through observations of near-surface flows by the Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI) (Löptien et al., 2018; Liang et al., 2019).

Standing Rossby waves, or r modes, are toroidal modes of oscillation of a rotating, fluid body for which the dominant restoring force against the pressure gradients is the Coriolis force (Lanza et al., 2019; Hathaway & Upton, 2021). Rossby waves are associated with an undulation of a flow resulting in a pattern of radial vorticity of alternating sign. They are understood to form in the high atmosphere on Earth, heavily influencing global weather. For the Sun we observe that Rossby waves propagate in the retrograde direction, in the Carrington reference frame.

Recently, Löptien et al. (2018) provided an unambiguous detection of sectoral solar r modes by tracking the horizontal flows of granules in the solar photosphere during a 6-year period, using observations by SDO/HMI. Following this study Liang et al. (2019) confirmed the detection of solar r modes with time-distance helioseismology to measure deeper, subsurface flows in the meridional direction along the solar equator using both Solar and Heliospheric Observatory Michelson Doppler Im-

further increased with a longer baseline, allowing for more accurate inferences on the SMMF morphology.

With more time on the project, it would also be useful to develop a technique similar to Kutsenko et al. (2017) and [Bose & Nagaraju \(2018\)](#), which allows the SMMF to be dissected into regions and features on the disc.

6 Rossby Modes in the Solar Mean Magnetic Field

6.1 Introduction

Rossby waves, as first derived by Rossby & Collaborators (1939), have recently been discovered in the Sun through observations of near-surface flows by the Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI) (Löptien et al., 2018; Liang et al., 2019).

Standing Rossby waves, or r modes, are toroidal modes of oscillation of a rotating, fluid body for which the dominant restoring force against the pressure gradients is the Coriolis force (Lanza et al., 2019; Hathaway & Upton, 2021). Rossby waves are associated with an undulation of a flow resulting in a pattern of radial vorticity of alternating sign. They are understood to form in the high atmosphere on Earth, heavily influencing global weather. For the Sun we observe that Rossby waves propagate in the retrograde direction, in the Carrington reference frame.

Recently, Löptien et al. (2018) provided an unambiguous detection of sectoral solar r modes by tracking the horizontal flows of granules in the solar photosphere during a 6-year period, using observations by SDO/HMI. Following this study Liang et al. (2019) confirmed the detection of solar r modes with time-distance helioseismology to measure deeper, subsurface flows in the meridional direction along the solar equator using both Solar and Heliospheric Observatory Michelson Doppler Im-

ager (SOHO/MDI) and SDO/HMI data, covering 21 years. In addition, Hanasoge & Mandal (2019) were also able to show the detection of solar r modes using a normal-mode coupling technique on 2 years of SDO/HMI data. In both of the observation conducted by Löptien et al. (2018) and Liang et al. (2019), the average lifetime of the r modes were on the order of several months, and even greater than a year for specific modes.

By monitoring the proper motions of solar super-granules using a local correlation tracking method Hathaway & Upton (2021) also report observing low latitude Rossby waves in full-disc Doppler images obtained by SDO/HMI, extending the measurements of Rossby waves to greater depths in the solar atmosphere, by an order of magnitude. The r modes observed using the super granules have lifetimes which are only slightly longer than the Carrington rotation period, hence in slight disagreement with Löptien et al. (2018) and Liang et al. (2019), which Hathaway & Upton (2021) claim may be due the waves getting in and out of phase with each other as the low wave number waves propagate faster than the higher wave number waves. As these observations are at low latitudes, it is possible that they are linked to active magnetic regions and hence could be manifested in other sources of magnetic data.

Here we present the search for evidence of a magnetic signature of the global r modes using over 20 years of Birmingham Solar Oscillations Network (BiSON) observations of the Line Of Sight (LOS) Solar Mean Magnetic Field (SMMF) (for details on the SMMF and BiSON, see Chaplin et al. (2003) and references within). *and Chapter 5 of this?*

In Chapter 5, the power spectrum of the BiSON SMMF observations was modelled using a combination of Lorentzian peaks and a convolution with the window function to account for the effect of gaps in the data. The full spectrum and the fit are shown in Figure 6.1.

We can divide through by the model to give a residual spectrum, in which we have searched for the magnetic signatures of r modes. In this work the residual spectrum

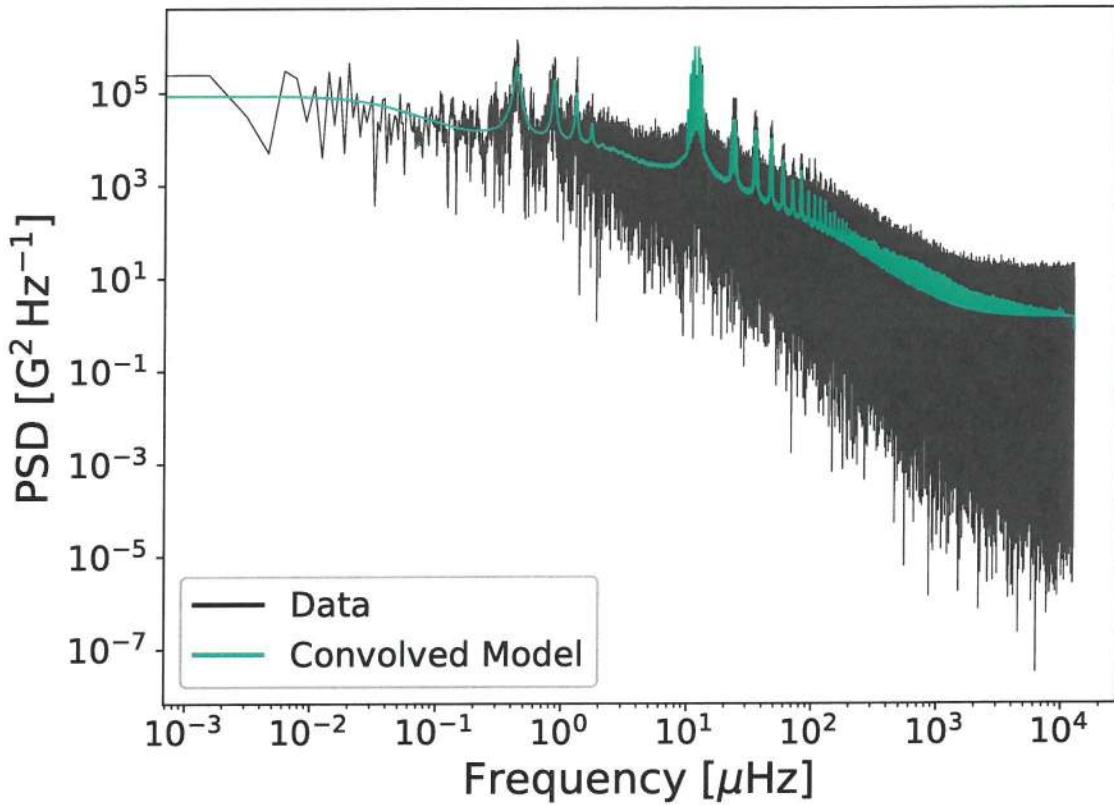


Figure 6.1: Full, modelled power spectrum of the BiSON SMMF on logarithmic axes. The data is displayed in black and the convolved model using symmetric Lorentzian peaks is shown in green.

was analysed to investigate the presence of r modes. Furthermore, analysis was performed using simulated data to better understand how an annual modulation of observations would affect any present r -mode frequencies.

6.2 Theory

The detailed theory of the effect of r modes on observational data was re-visited recently by Lanza et al. (2019) in light of the solar observations, in an effort to determine the effect of the r modes on radial velocity detections of exoplanets.

They showed that under the assumption of a slow, uniformly rotating sphere (with angular velocity, Ω , where $\Omega^2 \ll GM R^{-3}$), the frequencies of global r modes

in the Carrington rotating frame is well approximated by:

$$\nu_{carr} = -\frac{2m\Omega}{l(l+1)}, \quad (6.1)$$

where $l > 0$ is the angular degree and m is the azimuthal order (Löptien et al., 2018; Lanza et al., 2019):

In an inertial frame the observed r mode frequencies will be (Lanza et al., 2019):

$$\nu_{in}(l, m) \approx m\Omega - \frac{2m\Omega}{l(l+1)} = m\Omega \left(1 - \frac{2}{l(l+1)}\right), \quad (6.2)$$

where Ω is the mean sidereal rotation rate, and l and m are the angular and azimuthal degree, respectively. Sectoral Rossby waves are obtained by setting $l = m$ in this equation. A consequence is that they propagate with a retrograde phase velocity as $\nu/m = -2\Omega/[m(m+1)] < 0$.

An Earth-based observer, orbiting the Sun, shall expect to observe frequencies adjusted by the orbital frequency, $\nu_{\oplus} \approx 31.7\text{Hz}$, given by:

Space

$$\nu_{obs}(l, m) = \nu_{in}(l, m) - m\nu_{\oplus}. \quad (6.3)$$

In addition, due to the tilt of the ecliptic with respect to the solar equatorial plane (the solar B_0 angle), the visibility of the modes will vary on a timescale of 1 year, meaning we expect to actually observe split peaks at frequencies of $\nu_{obs}(l, m) \pm \nu_{\oplus}$ (Lanza et al., 2019).

Based on this theory, we can predict the frequencies at which to search for r modes. These frequencies are summarised in Table 6.1, along with the frequencies observed in other studies.

In Figure 6.2 we show the residuals spectrum with vertical lines indicating the locations of predicted r modes, based on Table 6.1. The solid lines indicate the location of the expected mode and the dashed lines show the expected locations of

Table 6.1: Predicted⁺ and observed[◦] r mode frequencies for combinations of l and m . Predicted frequencies and conversions of observations to different frames of reference use equation (6.1), equation (6.2), and equation (6.3), with $\Omega = 453.1$ nHz. The predicted splitting for the B_0 angle variation is also provided. The key for the source column is: LPT for Löptien et al. (2018), LNG for Liang et al. (2019), and LZA for Lanza et al. (2019).

Frequency	Source	$l = m = 2$	$l = m = 3$	$l = m = 4$	$l = m = 5$
ν_{carr} [nHz]	LPT [◦]	–	-259	-194	-157
	LNG [◦]	–	-253	-198	-156
	LZA ⁺	-302.1	-226.6	-181.2	-151.0
ν_{in} [nHz]	LPT	–	1100	1618	2109
	LNG	–	1106	1614	2110
	LZA	604.1	1132.8	1631.2	2114.5
ν_{obs} [nHz]	LPT	–	1005.2	1491.7	1950.1
	LNG	–	1011.2	1487.7	1951.1
	LZA	540.8	1037.7	1504.4	1956.0
$\nu_{obs} + \nu_{\oplus}$ [nHz]	LPT	–	1036.9	1523.3	1981.8
	LNG	–	1042.9	1519.3	1982.8
	LZA	572.4	1069.4	1536.1	1987.7
$\nu_{obs} - \nu_{\oplus}$ [nHz]	LPT	–	973.6	1460.0	1918.4
	LNG	–	979.6	1456.0	1919.4
	LZA	509.1	1006.0	1472.7	1924.3

modes if split by the B_0 variation.

There appear to be resolved peaks in the vicinity of the $l = m = 2$ r mode and perhaps also the upper B_0 -variation-modulated $l = m = 3$ r mode. We therefore tested for the presence of these modes in the residual spectrum.

6.3 Methodology

6.3.1 Testing the Residual Spectrum

In order to investigate the presence of Rossby wave modes in the power spectrum of the BiSON SMMF, statistical significance tests were employed using a false-alarm approach, to test the probability of finding prominent narrow-band power in the residual spectrum.

We assume negative exponential statistics (i.e. χ^2 2-degrees of freedom distribution), and that the bins in the power spectrum are uncorrelated. This was tested

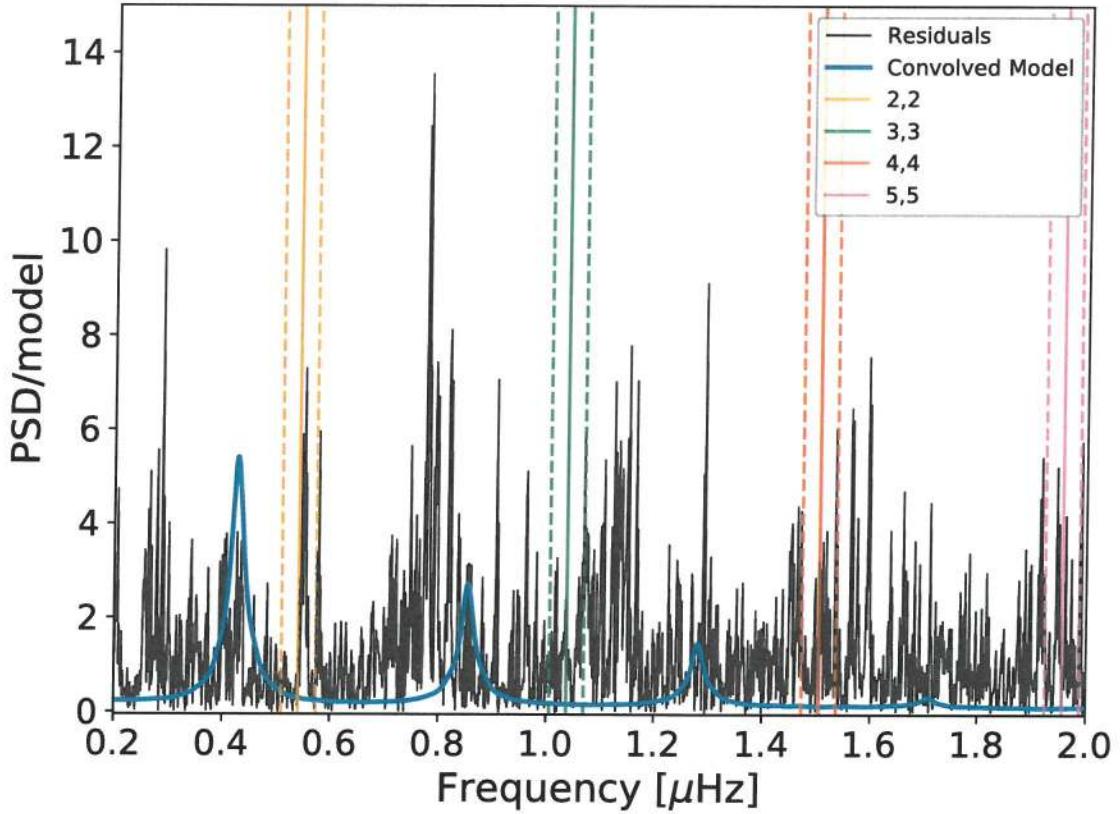


Figure 6.2: Residual power spectrum of the BiSON SMMF. Over plotted in the dark blue curve is the model of the main SMMF signal which was divided out from the raw spectrum. Also over plotted as vertical solid lines are the expected locations of the 4 lowest-frequency sectoral r modes and the dashed lines, the locations of the B_0 variation frequency splitting. Dashed lines represent ± 31.7 nHz (i.e. representing the frequencies of splitting due to the variation in the B_0 angle.)

using artificial data created with the same window function as the BiSON observations, and we showed that this assumption was suitable. We then calculated the false alarm probability, or probability to observe power purely by chance in a given frequency bin, ν , that is greater than or equal to a given threshold. The probability to observe power in a given frequency bin, ν , that is greater than or equal to $P(\nu)$ is:

$$p[P(\nu)] = \frac{1}{P_{lim}(\nu)} \exp\left(-\frac{P(\nu)}{P_{lim}(\nu)}\right) \text{ or; } p[P(\nu)] = \frac{1}{\langle P(\nu) \rangle} \exp\left(-\frac{P(\nu)}{\langle P(\nu) \rangle}\right), \quad (6.4)$$

where $P_{lim}(\nu)$ is the limit spectrum or $\langle P(\nu) \rangle$ is a well-fitting model/estimate of the spectrum. Considering a relative power approach (i.e. considering the power relative

to the mean level or model of the Power Spectral Density (PSD)), equation (6.4) may be written more concisely as:

$$p(s_\nu) = e^{(-s_\nu)}, \quad (6.5)$$

where,

$$s_\nu = P(\nu)/\langle P(\nu) \rangle, \quad (6.6)$$

and $\langle P(\nu) \rangle$ is reduced to 1 when we compare the power relative to the power spectrum residuals.

In reality, we used the χ^2 cumulative distribution function to compute the probability of power, which is given by equation (6.7), where k is the number of degrees of freedom, $\gamma(s, t)$ is the lower incomplete gamma function and $P(s, t)$ is the regularised gamma function:

$$F(x; k) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})} = P\left(\frac{k}{2}, \frac{x}{2}\right). \quad (6.7)$$

Using these expressions, we can rewrite the equation for $P(s_\nu)$ as given by equation (6.8):

$$p(s_\nu) = 1 - F(2s_\nu; k) = 1 - P\left(\frac{k}{2}, s_\nu\right). \quad (6.8)$$

This formulation provided the capability to compute the probability of statistically significant peaks in the residuals for re-binned data, for when data is binned over n -bins, the statistics becomes χ^2 with $2n$ degrees-of-freedom (Appourchaux, 2004).

The probability that a bin has power at or above the level s_ν is therefore given by equation (6.5), or more generally by equation (6.8), hence the probability that we fail to find a bin with power at or above the level s_ν is $1 - p(s_\nu)$; thus the probability of failing to find a bin with power at or above s_ν in N -bins in the spectrum is $[1 - p(s_\nu)]^N$. Therefore the probability to find at least one bin with power at or

above s_ν in N -bins in the spectrum is:

$$p_N = 1 - [1 - p(s_\nu)]^N, \quad (6.9)$$

where a low value for p_N indicates that the spike in power in that bin is unlikely to be a statistical fluctuation, and therefore is considered a statistically significant spike.

This can be generalised using the cumulative binomial distribution. The probability of finding at least r spikes in N -bins at or above the relative power level s_ν is given by equation (6.10), which is equal to equation (6.9) when $r = 1$,

$$p[r; p(s_\nu), N] = 1 - Pr(X < s_\nu) = \sum_{r=r}^N \binom{N}{r} p(s_\nu)^r [1 - p(s_\nu)]^{N-r}. \quad (6.10)$$

By applying equation (6.10) to the residuals of in the power spectrum, we can test whether there are any significant peaks in the residual power spectrum. Again, a low value for $p[r; p(s_\nu), N]$ indicates that the power in that bin is unlikely to be a statistical fluctuation.

6.3.2 Modelling r mode Profiles

In the location of any suspected r modes in the residual spectrum, we can model the profile of the peak by using a Lorentzian distribution using the form expressed by:

$$P(\nu) = \frac{2A^2/(\pi\Gamma)}{1 + (2(\nu - \nu_0)/\Gamma)^2}, \quad (6.11)$$

where A is the Root Mean Square (RMS) amplitude of the signal, Γ is the line-width of the distribution, and ν_0 is the central frequency of the distribution. This follows the methodology adopted by Löptien et al. (2018) and Liang et al. (2019).

Parameter estimation was performed in a Bayesian manner using a Markov Chain Monte Carlo (MCMC) fitting routine. The likelihood function used was the χ^2

degrees-of-freedom (Handberg & Campante, 2011; Davies et al., 2014). The prior information on each of the parameters used during the MCMC is discussed below, in Section 6.4.

Modelling was performed using the `pymc3` No U-Turn Sampler (NUTS) extension to a Hamiltonian Monte Carlo (HMC) sampling algorithm (Salvatier et al., 2016). Convergence was interrogated using the Gelman-Rubin \hat{R} diagnostic factor (Gelman & Rubin, 1992), using the criteria that chains did not converge if $\hat{R} > 1.01$.

6.4 Results

6.4.1 Testing the Residual Spectrum

The residual spectrum was shown in Figure 6.2, highlighting locations of r mode frequencies predicted by Lanza et al. (2019). It is clear from Figure 6.2 that there appears to be a resolved peak of narrow-band power in both the location of the $l = m = 2$ r mode and perhaps also the upper B_0 -variation-modulated $l = m = 3$ r mode.

The statistics tests were performed on the residual spectrum for various re-binning factors, n . The plots summarising the statistics tests are shown in Figure 6.3 for re-binning factors of $n = 1, 2, 5$, and 10.

These tests suggest that the feature around 0.5 μ Hz may be a signal related to r modes. This peak is particularly interesting despite the other potentially significant peaks near 0.75 μ Hz. The peaks near 0.75 μ Hz sit on top of the main rotational harmonics, and there is not a predicted r mode at those frequencies. The peak at 0.5 μ Hz does not sit on top of the main rotational harmonics and its location is roughly correct for the $l = m = 2$ mode. Furthermore, this peak consistently has a low False Alarm (FA) probability. In Figure 6.3b, there is compelling evidence to suggest that this feature is significant, in particular, as the peak is well-resolved in the second panel. In order to solidify this conjecture, we aimed to fit a model to the residuals around this peak in order to confirm whether the properties of the peak

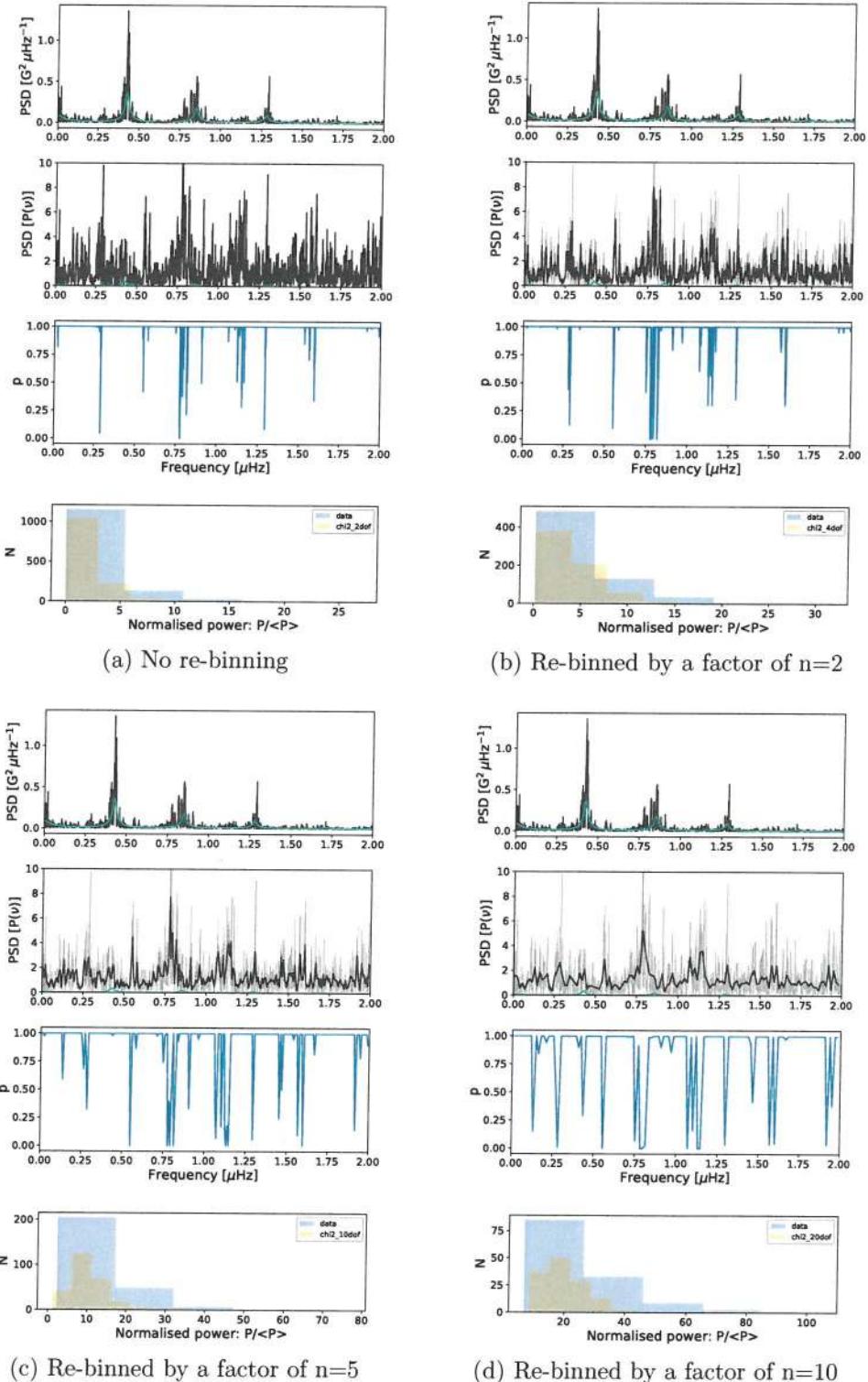


Figure 6.3: Realisations of the statistics tests on the BiSON data for different re-binning factors (n). The panels of each sub figure are: (top) the full PSD and fit, (second panel) the full and re-binned residuals, (third panel) the probability of statistical noise in each bin, (bottom) distribution of the residuals, blue, compared to a χ^2 $2n$ -DOF, orange.

resembled those suggested by Löptien et al. (2018), Liang et al. (2019), and Lanza et al. (2019).

6.4.2 Modelling r mode Profiles

Using the model for the Lorentzian peak (eq. 6.11), with an additional parameter (c) for additional noise, we modelled the spectrum around the location of the potential $l = m = 2$ mode. The parameter space was sampled using `pymc3` and uniform priors were used, providing reasonable boundaries on each parameter, as detailed below.

$$\nu_0 \sim \mathcal{U}(0.54, 0.56) \text{ } \mu\text{Hz}$$

$$\Gamma \sim \mathcal{U}(0.00, 0.05) \text{ } \mu\text{Hz}$$

$$A \sim \mathcal{U}(0.0, 500.0) \text{ mG}$$

$$c \sim \mathcal{U}(-0.5, 0.5) \text{ G}^2 \text{ Hz}^{-1}$$

The results of the fit are given in Table 6.2 and the fit to the residuals is shown in Figure 6.4.

Table 6.2: Median posterior values of the Lorentzian model for the r mode peak in the BiSON SMMF PSD. Numbers in brackets denote uncertainties on the last 2 digits, and all uncertainties correspond to the 68% credible intervals either side of the median.

Parameter	Value	Unit
ν_0	$0.5500^{(+19)}_{(-19)}$	μHz
Γ	$0.0052^{(+49)}_{(-28)}$	μHz
A	$27.1^{+7.9}_{-5.9}$	mG
c	$-0.002^{+0.059}_{-0.029}$	$\text{G}^2 \text{ Hz}^{-1}$

The median of the posterior distribution on the width parameter suggests an e -folding lifetime ($1/(\pi\Gamma)$) of around 710^{+670}_{-380} days ($\sim 1.9^{+1.8}_{-1.0}$ years), which follows the order of magnitude of the lifetime predicted and observed for low- m Rossby

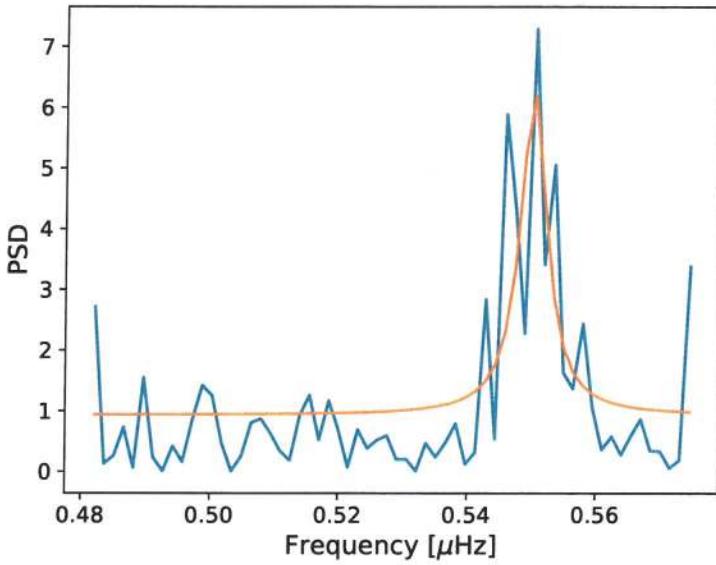


Figure 6.4: Model fit to the r mode in the BiSON PSD residuals using the median values from the posterior distributions of each parameter.

modes. The lowest r mode observed, $l = m = 3$, was shown to exhibit a lifetime of over a year (~ 1.4 years) by Liang et al. (2019). The $l = m = 4$ were observed by both Löptien et al. (2018) and Liang et al. (2019) to have a lifetime of ~ 0.6 years and ~ 0.3 years, respectively. There seems to be a slight increasing trend in the mode lifetimes observed by Löptien et al. (2018) and Liang et al. (2019), of longer lifetimes for lower m ; therefore, the long lifetime found here for the mode is entirely reasonable for the $l = m = 2$ mode and in-line with current observations.

The median value of the RMS amplitude of the peak was measured to be $27.1_{-5.9}^{+7.9}$ mG, which equates to a Radial Velocity (RV) amplitude of $\sim 7.8_{-1.7}^{+2.3}$ cm s $^{-1}$. Lanza et al. (2019) state that the maximum RV amplitude of the $l = m = 2$ mode is ~ 24.5 cm $^{-1}$, meaning our observed peak is around a third of the maximum RV amplitude one might have expected to observe. This is however an upper limit given by Lanza et al. (2019), and therefore the lower amplitude in the model should not be concerning.

In particular, however, we can see that the background is ~ 0 G 2 Hz $^{-1}$, which is expected. Due to the agreement between the central frequency of the fit to the $l = m = 2$ mode, the agreement in the order of magnitude of the e -folding lifetime,

and the amplitude of the mode residing below the upper limit suggested by Lanza et al. (2019), there is evidence to suggest that this peak, which has shown to be significant through the false-alarm statistics tests, could be the $l = m = 2$ sectoral r mode, observed in the BiSON SMMF data. We will cross-check this by further investigation using simulated data and, more directly, by comparison to other sources of SMMF observations, to determine whether the signal is present there too.

6.5 Discussion

Despite the results presented in the previous section there remained an open question on the way the Rossby waves would manifest themselves in the power spectrum. The BiSON power spectrum was also compared to the power spectra of the Wilcox Solar Observatory (WSO) SMMF and the SDO/HMI SMMF to cross-reference the finding.

6.5.1 Manifestation of Rossby Waves in the Power Spectrum

It was suggested by Lanza et al. (2019) that the mode should be split into two frequencies due to the annual variation of the B_0 angle, but the observed peak in the BiSON spectrum is located at approximately the location of the central frequency and is not split into significant peaks, separated by the predicted separation, due to this modulation. We needed to determine if this was physically observable.

Figure 6.5 shows a schematic diagram of the flow of a $l = 2 = m$ sectoral r mode. One can clearly see from the more visible purple region of the schematic, the Southern Hemisphere flow is oriented out of the page, whereas the Northern Hemisphere flow is oriented into the page. Due to the B_0 modulation, a varying of the sign of the flow would be observed over this region, i.e the velocity of the flow. In the more red-green regions of the schematic, the flow is more transverse, hence this would contribute less to the effect of the B_0 modulation.

We needed to therefore understand whether the SMMF observations have a hemispheric dependence that would lead to a sign change in the observations due to the B_0 variation. In addition, this raised the question of how the mode was affected in the power spectrum by the B_0 modulation; either split into separate peaks as suggested by Lanza et al. (2019) or instead was it possible that we could have a situation where the mode at the central frequency remained?

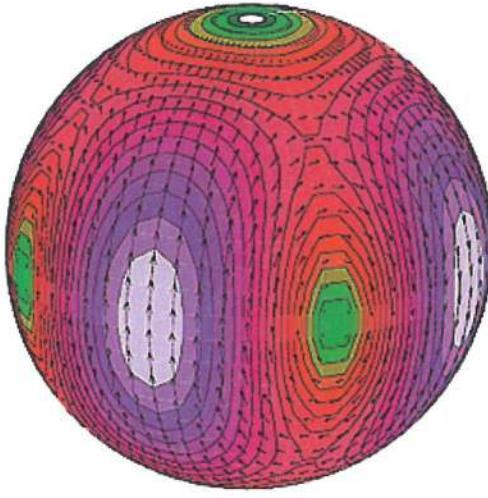


Figure 6.5: Mode displacement schematic for an $l = m = 2$ r mode (Strohmayer & Mahmoodifar, 2014)

To investigate the splitting of the mode, a simple model was produced whereby a sinusoidal function (with a period of ~ 25 days) was modulated by either a cosine or rectified cosine function (with a period of 1 year). In the former, using the cosine modulation, this represents observing the sign of the flow varying with the B_0 modulation, i.e. dubbed velocity modulation. Conversely in the latter simulation, this instead represents a variation of the amplitude, and it does not change the sign. It should be noted that the investigation of the mode splitting relies upon the assumption that the magnetic r mode signal has similar characteristics to the observed SMMF signal, i.e. dubbed amplitude modulation. Figure 6.6 shows the time series of the two cases to more clearly show their difference. The power spectrum of each case was then computed and these are shown in Figure 6.7 to demonstrate the differences between the modes produced.

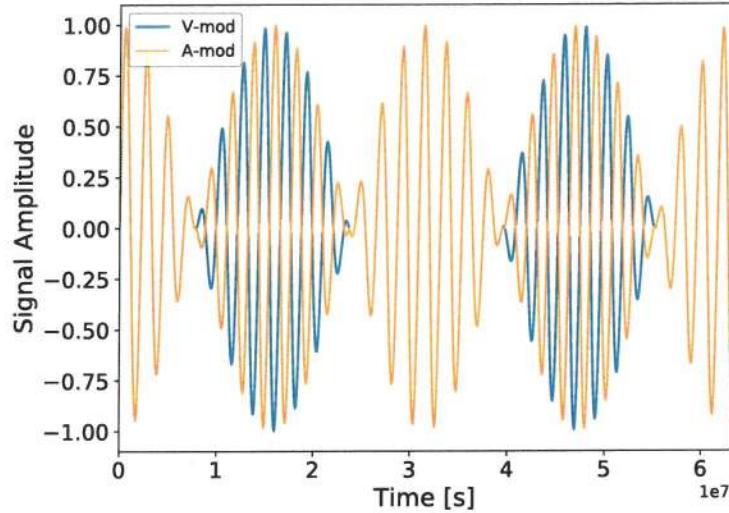


Figure 6.6: Time series of the velocity and amplitude modulation toy model simulations. The blue curve shows the velocity modulation, i.e. modulating using a cosine with period of 1 year, whereas the orange curve shows the amplitude modulation, i.e. modulating using a rectified cosine with period 1 year.

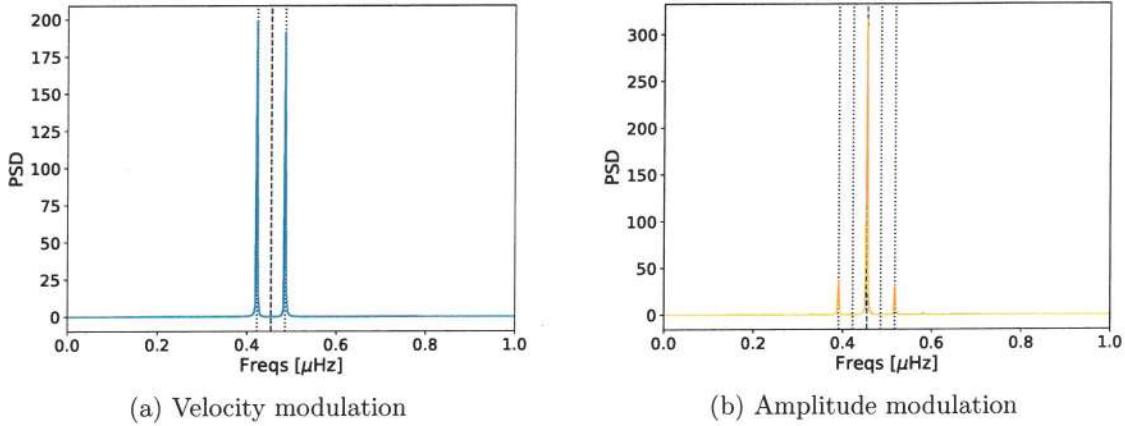


Figure 6.7: Power spectra for the two modulation methods, showing the difference in the way the modulation has changed the frequency of the observed mode. The dashed, vertical line shows the location of the central frequency and the dotted, vertical lines show multiples of ν_{\oplus} from the central frequency.

One can clearly see the difference in the power spectrum produced in each of the cases. In the velocity modulation case, we see a splitting of the oscillation mode into two peaks split around the expected mode frequency by $\pm\nu_{\oplus}$. There exists no more power at the central mode frequency in this case and is in agreement with the scenario suggested by Lanza et al. (2019). In the amplitude modulation case, we see sidebands at $\pm 2\nu_{\oplus}$; however, the expected mode frequency remains in this scenario

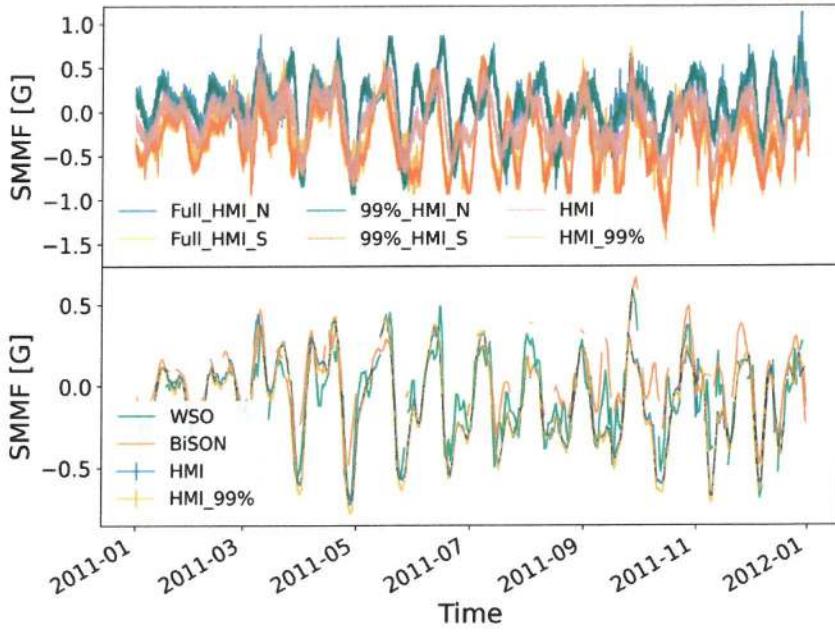
and has a significantly higher peak height, a ratio of 90:10 in favour of the central peak.

We have shown that it is possible to retain the central frequency of the r mode in the power spectrum if the B_0 modulates the amplitude of the observations and not the sign. With this known, it was then necessary to understand whether the two hemispheres of the Sun contribute signals that are more analogous with the velocity modulation or amplitude modulation. In the former, velocity modulation, we would expect to see a persistent anti-correlation between the two hemispheres. In the latter, amplitude modulation, we would expect to see the signals from each hemisphere that are correlated, which track each other and which can be both positive or negative.

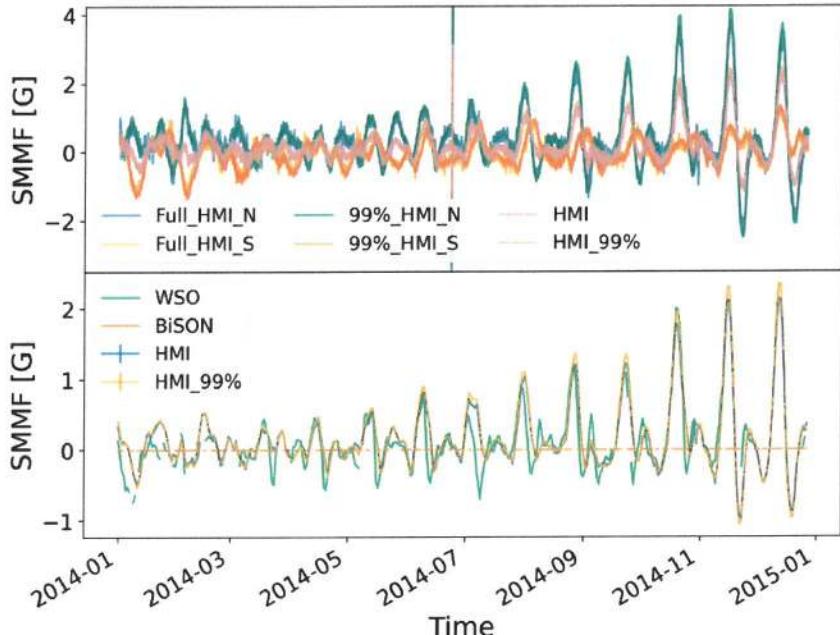
We investigated how the two hemispheres of the Sun contribute to the SMMF through analysis of SDO/HMI data. To do this, we acquired 720s-cadence magnetograms from SDO/HMI using the `SunPy` python module (Barnes et al., 2020) for the rising phase of solar cycle 24 during 2011, and for the maximum of cycle 24 during 2014. It was possible to separately average the Northern and Southern Hemispheres' contributions to the total, disc-averaged SMMF. The plot of this data is shown in Figure 6.8.

The hemispheric contributions to the total, disc-averaged, SMMF can be seen in Figure 6.8 to track each other during 2011, and both become positive or negative. By contrast, when observing the hemispheric contributions to the full SMMF in 2014, we see that there are frequent periods of strong anti-correlation between the Northern and Southern Hemispheres. There are also several periods in 2014 where the two hemispheres are correlated. This plot aids our understanding of how the r modes would manifest in the power spectrum due to the variation of the B_0 angle. Note that this assumes therefore that the r mode and SMMF signal behave in the same way.

As there are periods of both strong correlation and strong anti-correlation be-



(a) 2011



(b) 2014

Figure 6.8: SDO/HMI SMMF split into hemispheres and compared to other SMMF sources during (a) 2011, and (b) 2014. The top panel in each figure shows the north (N), south (S), and total disc-averaged mean magnetic field, for both the full solar disc and from pixels within 99% of solar radius. The bottom panels show a comparison between the SMMF, as observed BiSON, WSO, and SDO/HMI (full disc and the 99% disc).

tween the two hemispheres, this is a good indication that the r mode signal would result in a central frequency with sidebands due to the correlation between North

and South. But due to the existence of periods of anti-correlation between North and South, we also expect there to be some degree of frequency splitting in the power spectrum, but this is dependent on how prevalent the anti-correlation is over the entire solar cycle; in these short epochs however, we expect this to be minimal. We were therefore optimistic that this was the $l = 2 = m$ r mode.

As a further point, we see that there is both strong correlation and strong anti-correlation between the Northern and Southern Hemispheres; however, this does not necessarily mean that the r mode signal would directly manifest itself in the same way. We can see from Figure 6.5 that if the r mode observations are constrained to active latitude bands, closer to the equator, then the effects of the B_0 are less prominent.

6.5.2 Rossby Modes in Other Sources of SMMF Data

To further investigate whether the observation of the $l = 2 = m$ r mode is real, a comparison was made between the power spectrum of the BiSON observations of the SMMF and those from WSO and SDO/HMI, to determine whether the suspected r mode is visible in the other data. In the case of the WSO, the power spectrum was computed over the same observing epoch as the BiSON data (i.e. from 1992 – 2012); however, SDO was not launched until 2010, so for HMI the power spectrum was computed on data from 2010 – 2020, hence at approximately half the frequency resolution of WSO and BiSON.

Figure 6.9 shows the comparison of the WSO and SDO/HMI power spectra reflected around the x-axis against the BiSON power spectrum. In both cases we see a good agreement between the different sets of data on the location of the rotational mode in the SMMF, but there does not appear to be a visible r mode candidate in the WSO or the SDO/HMI data.

This provides concerning evidence that we may not, in fact, have observed the $l = 2 = m$ r mode in the BiSON power spectrum, and perhaps instead we have a

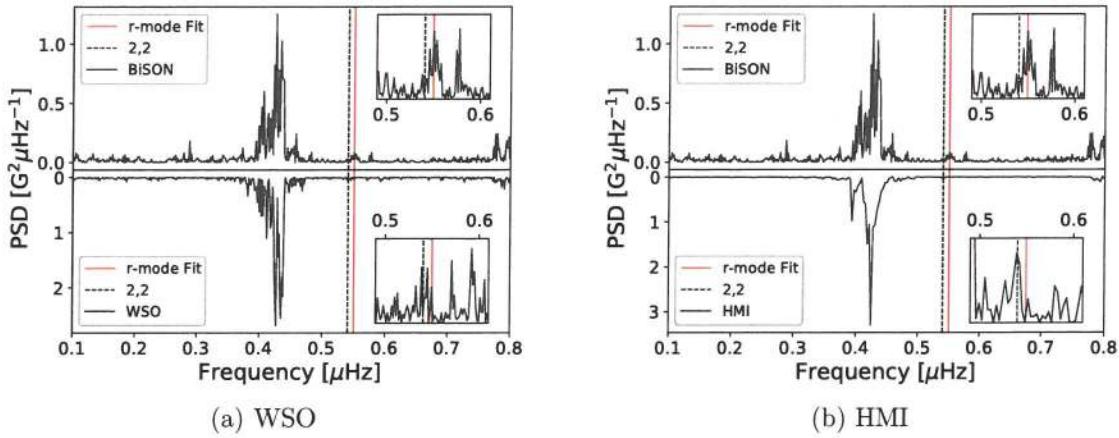


Figure 6.9: Comparison of the power spectra for BiSON, WSO, and SDO/HMI. In both figures, the top panel shows the BiSON PSD and the bottom panel shows either the WSO or HMI PSD. The dashed, black line shows the location of the theoretical $l = 2 = m_r$ mode frequency, and the red, solid line shows the location of the peak fit in the BiSON PSD residuals.

persistent noise source in the BiSON SMMF.

However compelling the results are, suggesting that we may have observed the $l = 2 = m_r$ mode in the BiSON power spectrum, it was impossible to ignore that it was not present in two other sources of SMMF observations – of which one of these telescopes was responsible for providing the recent observations of r modes documented in the literature (SDO/HMI). Owing to this, we cannot conclude that we have observed the $l = 2 = m_r$ mode in the BiSON power spectrum. We recommend that searching for r modes in the power spectra of the SMMF observations should be revisited in a few years, or say a solar cycle’s time, when the frequency resolution in each power spectrum has increased significantly, to provide a more insightful follow-up study.

6.6 Conclusion

After removing the model for the BiSON SMMF power spectrum we investigated the residual spectrum to search for evidence of a magnetic signature of global Rossby modes (r modes). A statistically significant peak was identified near the $l = 2 = m_r$ mode frequency calculated by Lanza et al. (2019), which was further modelled to

examine whether it was the r mode.

Using a model for the peak, following the description of r mode given by Löptien et al. (2018) and Liang et al. (2019), we identified the properties of the peak and compared it to the expected values in the literature.

To further interrogate the inferences on the statistically significant peak in the BiSON SMMF power spectrum, we used simple simulations to determine how the r mode may manifest itself in the power spectrum, and SDO/HMI hemispheric data were examined as a cross-reference.

As a final check, we compared the power spectrum of the SMMF observations from BiSON to those from WSO and SDO/HMI. There were no statistically significant peaks in the vicinity of the $l = 2 = m$ r mode in either of the other power spectra. Because this peak was only observed in one of three data sets, and particularly not in the SDO/HMI data which recent observations of sectoral Rossby waves in the Sun have all used, we could not conclude that the statistically significant peak in the BiSON SMMF was the $l = 2 = m$ r mode.

We leave the reader with the following points:

1. Through a series of false-alarm probability statistical tests, we have shown that there exists a statistically significant peak in the BiSON SMMF residuals spectrum which is located near the theoretical frequency of the $l = 2 = m$ r mode.
2. By modelling the peak as a Lorentzian profile we find that the peak has a central frequency of 550 ± 19 nHz (i.e. located ~ 9.2 nHz from the theoretical frequency), a line-width of $5.2_{-2.8}^{+4.9}$ nHz, and an amplitude of $\sim 27.1_{-5.9}^{+7.9}$ mG. This profile is within the upper limit for the amplitude of the $l = 2 = m$ r mode and the life time implied by the line-width is on the order of 1–2 years, which is in agreement with the observations by Löptien et al. (2018) and Liang et al. (2019).

3. Through the analysis of simulated data and hemispheric observations of the SMMF, we have shown that we should expect to see a prominent mode at the theoretical frequency, and not a split mode due to the effect of the B_0 variation, which supported the investigation that the peak may be the $l = 2 = m r$ mode.
4. By comparing the power spectrum of the SMMF observed by BiSON, to those of WSO and SDO/HMI, we have shown that the observed, statistically significant peak is not manifested in either the WSO or SDO/HMI spectra, therefore ruling it highly unlikely that the observed peak in the BiSON spectrum is the $l = 2 = m r$ mode.

As we collect more observations of the SMMF using BiSON, the frequency resolution of the power spectrum increases. An obvious next step in this work is to collect more observations of the SMMF with BiSON, to further investigate if this suspected mode remains resolved, or whether it diminishes into the noise.

7 Conclusions and Future Prospects

In this thesis, three studies have been presented, which investigated: using Cosmic Ray (CR) detectors for space weather applications, the impact of solar activity on Galactic Cosmic Rays (GCRs), and solar interior-atmosphere linkages using observations the Solar Mean Magnetic Field (SMMF).

In Chapter 2 we explored the properties and observations of the High School Project on Astrophysics and Research with Cosmics (HiSPARC) experiment, to determine the feasibility of its use for monitoring space weather events. Using simulations of the interactions between CRs and the Earth's magnetosphere, we were able to calculate the rigidity cut-off and Asymptotic Viewing Directions (AVDs) of the HiSPARC stations. We showed that the rigidity cut-off limits the observable Primary Cosmic Rays (PCRs) to those with energies on the order of and above $\sim 10^9$ eV. This highlighted that HiSPARC stations would not observe the particles most susceptible to space weather events, i.e. those with energies $\sim 10^7 - 10^9$ eV.

We showed that we were unable to clearly detect signatures, in the raw HiSPARC data, of the Forbush Decreases (FDs) or Ground Level Enhancements (GLEs) that occurred over the lifetime of the HiSPARC network. In addition, we showed that the effects of meteorological conditions caused variations in the data, which limited our ability to detect these events. A method to correct for the effects of atmospheric pressure and temperature was successfully demonstrated and applied to the HiSPARC data. Following the correction of the atmospheric effects, the search for evidence of GLEs was repeated with the corrected HiSPARC data. However, we con-

cluded that we were also unable to clearly claim any observations of space weather events in the corrected HiSPARC data.

Air Shower (AS) simulations were employed to calculate the flux of muons at ground level. For low-energy CRs ($\sim 10^9$ eV), we found the flux was small, producing very diffuse air showers of only a few muons, instead of the Extensive Air Showers (EASs) that HiSPARC intends to observe. We showed the configuration of the HiSPARC stations, which strongly relies on the triggering of multiple detectors within a station, biased observations to higher energy PCRs, hence limiting the capabilities of using HiSPARC network for space weather observations. This provided some explanation of why we were unable to detect the space weather events in the HiSPARC data.

Furthermore, we ran simulations to predict the increase in the HiSPARC count rate for some of the largest GLEs to-date. This showed that, on average, we expect an increase in the muon count rate using the HiSPARC detectors of $< 1\%$, for a ‘typical’ GLE. It was shown that only the most energetic events, with a lower occurrence rate, would induce an increase in the HiSPARC counts by $\gtrsim 5\%$ and hence be potentially detectable. This provided further evidence to explain why we were unable to observe the GLEs and FDs in the HiSPARC data. Therefore, we concluded that the HiSPARC network is generally incompatible with monitoring the lower limits of space weather activity, and only suitable as a monitor of the rarer and more extreme events.

Leading on from these results, in Chapter 3 we presented an alternative HiSPARC station configuration, with a novel arrangement of the detectors, and investigated its performance for use monitoring space weather events. Firstly, we outlined the configuration and technical set-up of the station; secondly, we performed the relevant atmospheric corrections, where we showed that the new station provides an accurate measure of the temperature inside the roof boxes for station 14008. Here, we also demonstrated that these temperature data were also an accurate measure

I'm sure I didn't
catch all of these!

for the detectors of station 14001, which are located on the same building, on the University of Birmingham campus.

Using a Bayesian method to sample from the posterior distribution, we found the mean count rate of the new station configuration was $\sim 80 \mu\text{s}^{-1}$, which was in good agreement with the predicted values from the air shower simulations in Chapter 2. Furthermore, we determined the noise from spurious counts is of about $0.0043 \pm 0.0002 \mu\text{s}^{-1}$, which is negligible compared to the Poisson noise representing $\sim 11\%$ of the signal. Comparing the data to that collected by a nearby Neutron Monitor (NM) station, Dourbes—in Belgium, we showed that there was a good visual agreement between the two data sets. This relationship should be continually monitored, as it will be instrumental in the verification of the new station configuration when a space weather event occurs in the next Solar Cycle.

Simulations of artificial data were performed to assess the likelihood of observing GLEs in the new station configuration. We demonstrated that with 10-s cadence observations we expect to be able to detect GLEs with a magnitude of $\gtrsim 3 - 4\%$. Furthermore, through averaging the data into 1- or 5-minute bins, we showed this improved the sensitivity to observe GLEs with magnitudes $\gtrsim 1.5 - 2.0\%$. These values are in line with some of the predicted GLE magnitudes from Chapter 2, providing compelling evidence to suggest that we should be capable of observing GLEs in this configuration.

We also simulated the performance of a network of detectors in this configuration and showed that we can improve the sensitivity to observe GLEs with magnitudes on the order of $\sim 1\%$, through analysing the cross-correlation of nearby stations. However, we note that there is a strong dependence on decay time of the GLEs. We concluded that any upgrades to form a network of stations in this configuration should ideally use at least 5 stations, and 10 stations would be more beneficial.

In Chapter 4 we studied long-term variations of GCR intensity in relation to the

Sun Spot Number (SSN) during the most recent solar cycles. We investigated the time lag between the GCR intensity and the SSN, and the hysteresis effect of the GCR count rate against SSN for Solar Cycles 20–24.

We showed that in cycle 24, the GCR intensity lagged behind the SSN by 2–4 months, which was slightly longer than the preceding even-numbered solar activity cycles (approx. 0–1 months). We showed the lag was not as large as the preceding odd-numbered cycles, and cycle 24 followed the trend of a short or near-zero lag for even-numbered cycles. We concluded that the cause of the extended lag in cycle 24 compared to previous even-numbered cycles was related to the deep, extended minimum between cycle 23 and 24, and the low maximum activity of cycle 24.

In addition, we showed the difference in the shapes of the hysteresis plots for odd-numbered and even-numbered cycles. The hysteresis plots were modelled using both a simple linear model and an ellipse model; the results showed that cycle 24 followed the same trend as preceding even-numbered cycles and was best represented by a straight line rather than an ellipse.

The time lag analysis was repeated using data from HiSPARC station 501 (Nikhef). However, it was quantitatively concluded that there exists no correlation between the SSN and the muon count rate measured by HiSPARC station 501. The limiting factor to observe the effect was changes in set-up of the HiSPARC station over time, which counteracted the expected variation due to solar activity.

In Chapter 5, a frequency-domain analysis of over 20 years of high-cadence Birmingham Solar Oscillations Network (BiSON) observations of the SMMF was presented. If we convert a time series of the SMMF to the frequency-domain, a strong Rotationally Modulated (RM) signal appears as a series of peaks. This characteristic demonstrates that the source of the SMMF is long-lived, over several rotations. The power spectrum of the BiSON SMMF data was modelled to draw conclusions about the morphology of the SMMF, particularly focusing on the source of the rotationally

modulated component in the signal.

The duty cycle for the 40-second cadence observations was very low, hence the effect of the low fill on the power spectrum of the SMMF was investigated to inform how to best model the complete power spectrum. This highlighted that although there appeared to exist a red-noise-like, stochastic background component in the power spectrum, this was a feature originating from power aliasing, due to the low duty cycle of the observations. We had to be very cautious in our approach when modelling the power spectrum to ensure that Parseval's theorem was obeyed and that the effects of the window function were robustly accounted for.

Using a Bayesian approach, a model was fitted to the power spectrum. We found that the RM component had a frequency of $0.4270 \pm 0.0018 \mu\text{Hz}$. This frequency allowed us to infer the sidereal period of the RM signal to be 25.23 ± 0.11 days which suggested cycle-averaged latitude of $\sim 12^\circ$, thus linking the source to active bands of latitude on the Sun. From the width of the RM component peak, we were able to determine the lifetime of its source. We measured the lifetime to be 139.6 ± 18.5 days, which is in the region of $\sim 20 \pm 3$ weeks.

The measured properties of the RM component of the SMMF were consistent with Active Regions (ARs). The literature provided compelling arguments to suggest that sunspots were not the origin of the SMMF, therefore we concluded that, more generally, ARs and Magnetic Flux Concentrations (MFCs) are the source of the dominant, rotation signal in the SMMF, that are long-lived on the solar disc and exist in active latitudes. In addition, we demonstrated, numerically and analytically, that our ability to determine the linewidth and hence lifetime of the RM modes was unaffected by AR migration and differential rotation.

In Chapter 6 we further investigated the BiSON SMMF data to search for evidence of a magnetic signature of global Rossby modes (r modes) in the residual power spectrum. A well-resolved peak was identified near the predicted $l = 2 = m$ r mode frequency. Using a Bayesian modelling technique we found the peak was

centred of frequency of 550 ± 19 nHz (i.e. ~ 9.2 nHz from the predicted frequency, but within measured uncertainty). In addition we measured the width of the peak to be $5.2_{-2.8}^{+4.9}$ nHz, and amplitude to be $\sim 27.1_{-5.9}^{+7.9}$ mG. The properties of the measured peak were in agreement with observations of other sectoral r modes and in-line with the predictions for the $l = 2 = m$ r mode.

To understand the way the r mode would manifest in the power spectrum, due to the variation in the B_0 angle we generated simulated data and acquired additional hemispheric observations of the SMMF using full-disc magnetograms. Through the analysis of these data, we showed how one might expect to see a prominent mode at the theoretical frequency, and not a split mode due to the effect of the B_0 variation. This further supported the hypothesis that the peak may have been the $l = 2 = m$ r mode, assuming the magnetic r mode signal had similar characteristics to the observed SMMF signal.

Finally, we investigated the power spectra of the SMMF observed with the Wilcox Solar Observatory (WSO) and Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI), to verify if the peak was consistent across all observations. However, these observations showed no statistically significant peak in the location of the predicted $l = 2 = m$ r mode frequency. It therefore made it highly unlikely that the candidate peak in the BiSON spectrum was the $l = 2 = m$ r mode. Because this peak was only observed in one of three data sets, and particularly not in the SDO/HMI data which recent observations of sectoral Rossby waves in the Sun all used, we could not conclude that the candidate peak in the BiSON spectrum was the $l = 2 = m$ r mode.

Future Prospects

The first studies, presented in Chapter 2 and Chapter 3, of this thesis represented the beginning of exploring how the HiSPARC network can be used to monitor space weather. The results from the first chapter showed that the HiSPARC stations,

in their original configurations, are not suitable for observing the effects of space weather; however, unfortunately, there have been few space weather events that have occurred over the lifetime of the HiSPARC experiment and it is possible with more observations that we will observe a first space weather event with the existing HiSPARC network. Continued monitoring of the data is imperative to check whether future space weather events are measured. This should be further supported by the observations from the Global Neutron Monitor Network (GNMN) (Mishev & Usoskin, 2020), as they are proven to observe these events.

The air shower analysis to predict the variation in the muon count rate during space weather events was far from complete. We predicted the increase in muon intensity for only 6 of the 72 GLEs to-date. This was the limit of the analysis using the Model for Atmospheric Ionising Radiation Effects (MAIRE) tool used in this work. A more comprehensive analysis could be performed, using other sources of CR spectra during space weather events, to predict the effect on the muon flux of the other GLEs and also FDs.

The new station configuration shows promise to improve the space weather capabilities of the HiSPARC network and it could be the beginning of a network-wide re-configuration. The data collected by HiSPARC station 14008 should be maintained until at least 2026, to ensure a complete study is performed up to the maximum of Solar Cycle 25 (McIntosh et al., 2020; Pesnell, 2020), when it is more probable that space weather events will occur. However, it would be of significant benefit to support this configuration for as long as possible, to compare the performance against the original HiSPARC configuration for detecting space weather events.

The work presented in this chapter also investigated the effects of building a network of stations using this new configuration. It would be timely to create this network in the near-term, before cycle 25 maximum, to benefit from the improved sensitivity at the time when more space weather events are expected.

It is also of interest to revisit Ross & Chaplin (2019) (i.e. the work presented in Chapter 4) to: (i) confirm the conclusions are the same up to the end of cycle 24/start of cycle 25; (ii) investigate the properties of cycle 25; (iii) analyse consistent, good-quality data from the HiSPARC network, in both the original configuration and station 14008 configuration, to observe the solar cycle modulation of GCRs.

It would be highly desirable to improve the magnetic field capabilities of BiSON to increase the duty cycle of the data and reduce the effect of the window function. In addition, with more observations, the frequency resolution will improve, allowing for more accurate inferences on the SMMF morphology. Plans were in place to re-acquire observations of the SMMF using the Sutherland node of BiSON before the COVID-19 pandemic struck, which delayed the work. This will be done when it is possible to travel again. It would also be advantageous to provide this capability elsewhere in the network.

Finally, on the future prospects of the suspected Rossby mode in the BiSON data. Again, as we collect more observations of the SMMF using BiSON, the frequency resolution of the power spectrum improves. An obvious next step in this work is to collect more observations of the SMMF with BiSON, to further investigate if this suspected mode remains resolved, or whether it diminishes into the noise. If still resolved, the investigation should be repeated, to determine the source of the signal.

A Simulations of Cosmic Ray Air Showers using CORSIKA

There are several user-definable settings within Cosmic Ray Simulations for Kascade (CORSIKA). These settings are explained in-depth in the CORSIKA user's guide (Heck & Pierog, 2017). Here, the options selected when building and using CORSIKA are outlined.

For high energy, inelastic hadronic interactions within CORSIKA the QGSJET-II (Ostapchenko, 2006) model was selected. Interactions of hadrons with energies below 80 GeV are simulated using GHEISHA (Fesefeldt, 1985), which allowed for the simulation of Primary Cosmic Rays (PCRs) in the regime of Solar Cosmic Rays (SCRs). In addition to these hadronic interactions, electromagnetic interactions within the CORSIKA simulations were described by the EGS4 (Nelson et al., 1985) model. Furthermore CORSIKA has a minimum muon energy limit that can be simulated of 10 MeV. This limit is sufficient, as the lowest energy muons that are observable with HiSPARC are those with energy of ~ 4 MeV.

Simulation thinning was enabled to reduce the computation time of the simulations and reduce the output file size. The observation level at which point the simulation cease was set at 100 m above sea level (compared to the ~ 50 m typical of the stations; however, this difference is negligible for the air shower development). The pre-defined central European atmosphere in October was used for all simulations, and western-European magnetic field was used as calculated with the *Geomag*

programme (BGS, 2020): $B_x = 18.799 \mu\text{T}$ and $B_z = 44.980 \mu\text{T}$.

Proton and α -particle initiated air showers were generated with energies ranging from 10^9 to 10^{20} eV, and 4×10^9 to 10^{20} eV, respectively. In total ~ 230000 proton-initiated showers were simulated and ~ 180000 α -particle-initiated air showers were simulated. The simulated air-shower-initiating PCRs are listed Table A.1 (proton) and Table A.2 (α -particles).

Table A.1: Details of the proton-initiated air showers simulated using CORSIKA.

E_{PCR} (eV)	N_{sims}	E_{PCR} (eV)	N_{sims}
1.00E+09	10000	2.98E+12	1000
1.27E+09	10000	3.79E+12	1000
1.62E+09	10000	4.83E+12	1000
2.07E+09	10000	6.16E+12	1000
2.64E+09	10000	7.85E+12	1000
3.36E+09	10000	1.00E+13	1000
4.28E+09	10000	1.78E+13	100
5.46E+09	10000	3.16E+13	100
6.95E+09	10000	5.62E+13	100
8.86E+09	10000	1.00E+14	100
1.00E+10	10000	1.78E+14	50
1.13E+10	10000	3.16E+14	50
1.44E+10	10000	5.62E+14	50
1.83E+10	10000	1.00E+15	10
2.34E+10	10000	1.78E+15	10
2.98E+10	10000	3.16E+15	10
3.79E+10	10000	5.62E+15	10
4.83E+10	10000	1.00E+16	10
6.16E+10	10000	1.78E+16	10
7.85E+10	10000	3.16E+16	10
1.00E+11	10000	5.62E+16	10
1.27E+11	1000	1.00E+17	10
1.62E+11	1000	1.78E+17	10
2.07E+11	1000	3.16E+17	10
2.64E+11	1000	5.62E+17	10
3.36E+11	1000	1.00E+18	10
4.28E+11	1000	1.78E+18	10
5.46E+11	1000	3.16E+18	10
6.95E+11	1000	5.62E+18	10
8.86E+11	1000	1.00E+19	10
1.13E+12	1000	1.78E+19	10
1.44E+12	1000	3.16E+19	10
1.83E+12	1000	5.62E+19	10
2.34E+12	1000	1.00E+20	10

Table A.2: Details of the α -particle-initiated air showers simulated using CORSIKA.

E_{PCR} (eV)	N_{sims}	E_{PCR} (eV)	N_{sims}
4.00E+09	10000	1.00E+13	1000
4.28E+09	10000	1.78E+13	100
5.46E+09	10000	3.16E+13	100
6.95E+09	10000	5.62E+13	100
8.86E+09	10000	1.00E+14	100
1.00E+10	10000	1.78E+14	50
1.13E+10	10000	3.16E+14	50
1.44E+10	10000	5.62E+14	50
1.83E+10	10000	1.00E+15	10
2.34E+10	10000	1.78E+15	10
2.98E+10	10000	3.16E+15	10
3.79E+10	10000	5.62E+15	10
4.83E+10	10000	1.00E+16	10
6.16E+10	10000	1.78E+16	10
7.85E+10	10000	3.16E+16	10
1.00E+11	10000	5.62E+16	10
1.27E+11	1000	1.00E+17	10
1.62E+11	1000	1.78E+17	10
2.07E+11	1000	3.16E+17	10
2.64E+11	1000	5.62E+17	10
3.36E+11	1000	1.00E+18	10
4.28E+11	1000	1.78E+18	10
5.46E+11	1000	3.16E+18	10
6.95E+11	1000	5.62E+18	10
8.86E+11	1000	1.00E+19	10
1.13E+12	1000	1.78E+19	10
1.44E+12	1000	3.16E+19	10
1.83E+12	1000	5.62E+19	10
2.34E+12	1000	1.00E+20	10
2.98E+12	1000		
3.79E+12	1000		
4.83E+12	1000		
6.16E+12	1000		
7.85E+12	1000		

B Simulations of the Ground Level Enhancements for the Analysis of HiSPARC Station 14008

B.1 Model

The artificial data used in the simulations of the Ground Level Enhancements (GLEs), during the investigation of the performance of High School Project on Astrophysics and Research with Cosmics (HiSPARC) station 14008, were created using a very simple model which was physically motivated by GLE observations (Strauss et al., 2017).

There were several variables that allowed us to change the physics of the simulations. These were:

- T_d : Observing window
- t_0 : Time of GLE peak
- λ : Background count level
- σ : Detector noise level
- A : Percentage amplitude of GLE increase
- τ_r : Rise time of GLE

- τ_d : Decay time of GLE

Using these variables, the mathematical form of the model was expressed by:

$$y(t) = \sigma + \lambda(1+A) \cdot \begin{cases} \exp\left[\frac{t-t_0}{\tau_r}\right] & \text{if } t \leq t_0 \\ \exp\left[-\frac{t-t_0}{\tau_d}\right] & \text{if } t > t_0 \end{cases}, \quad (\text{B.1})$$

is that an 1dot? It should be cdot.

where t is the time in the range from $0 - T_d$ in 10-second intervals, to be representative of the real HiSPARC station 14008 data.

An example time series of a single GLE is shown in Figure B.1. This shows both the model used to generate the simulated data (black line), as well as the background level and noise with and without the GLE injected, blue and orange lines, respectively.

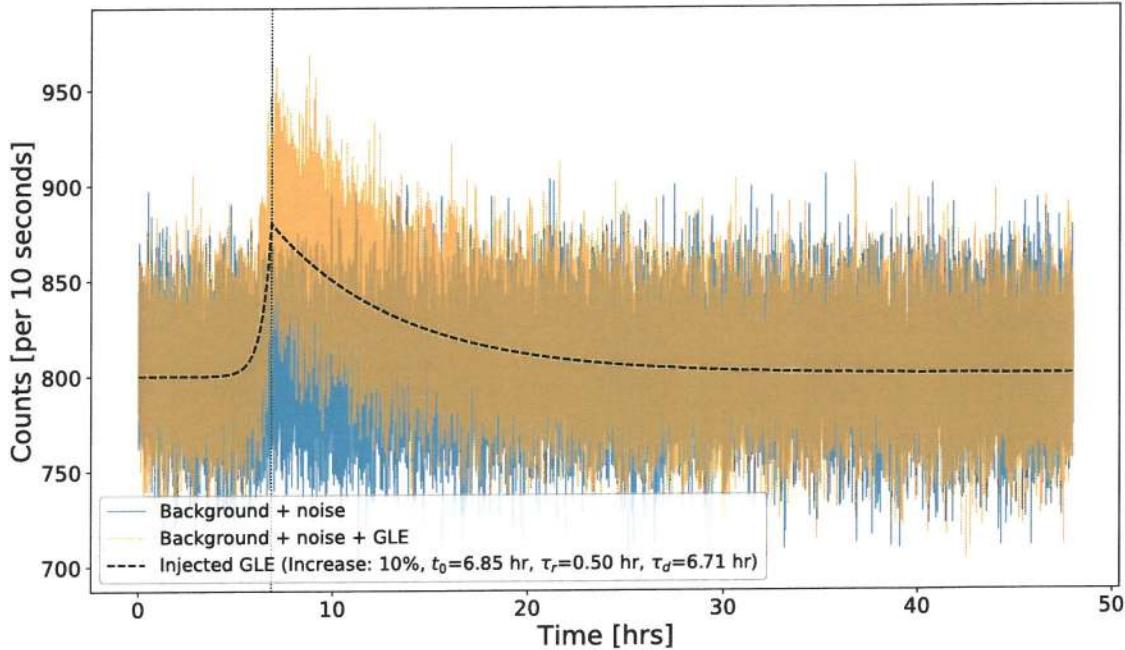


Figure B.1: Example of a realisation of the simulations, to generate an artificial time series with a 10% GLE injected. The black line shows the model for the GLE. The blue line shows the realisation of the signal without a GLE injected, and the orange line shows the realisation of the model with the GLE injected.

B.2 Configuration of the Simulations

The simulated data were generated using a Python script. A flowchart describing the steps in the simulation is shown in Figure B.2.

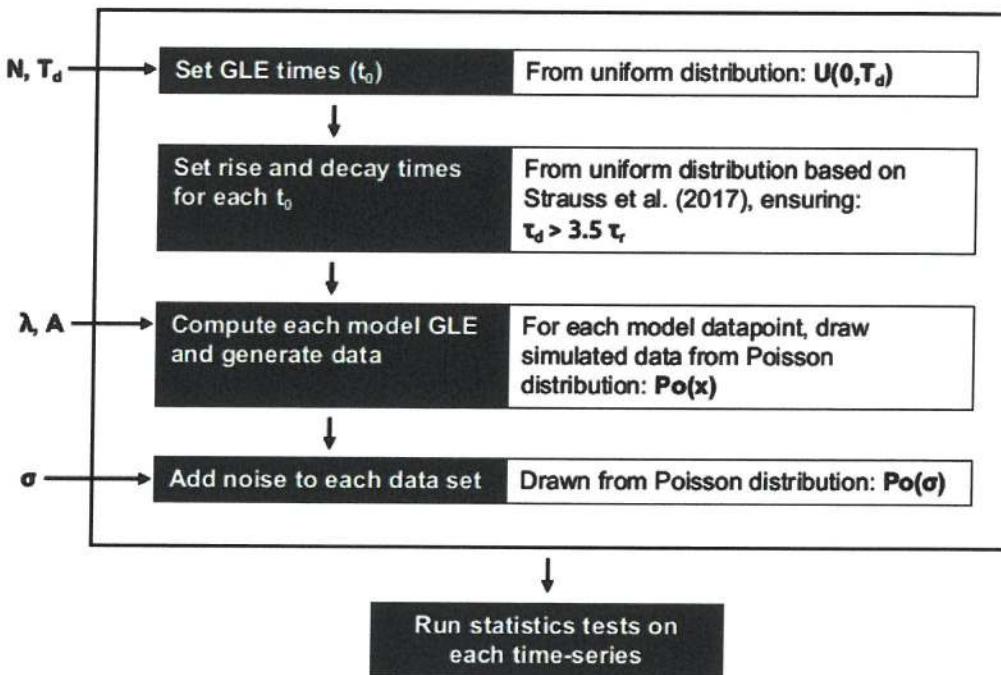


Figure B.2: Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.

The simulations required the user to input the number of iterations, N , to run. For each iteration, the time of the GLE peak, t_0 , was drawn from a uniform distribution between 0 and T_d . Rise and decay times were drawn from uniform distributions, $\mathcal{U}(0, 75)$ minutes and $\mathcal{U}(15, 500)$ minutes, based on the findings on the shape of GLE signals (Strauss et al., 2017). In addition we also ensured $\tau_d \gtrsim 3.5 \tau_r$, (Strauss et al., 2017).

With these properties drawn, and the user inputting the background level (λ), noise-level (σ), and percentage increase of the GLE (A), each individual GLE was simulated according to equation (B.1). After the creation of the simulated data, statistical tests were performed to determine whether the GLE is observed.

The above described the steps for a single station simulation. When multiple sta-

tions are simulated (i.e. n stations), the process is the same; however, it is repeated for each station, n_i . Another difference is that for each station the background level (λ), noise-level (σ), and percentage increase of the GLE (A) are drawn from narrow normal distributions, to simulate slight differences in the observations of the stations. Using the input values λ , σ , and A , we draw new values for each station: $\lambda \sim \lambda \times \mathcal{U}(0, 0.01)$, $\sigma \sim \sigma \times \mathcal{U}(0, 0.01)$, and $A \sim A \times \mathcal{U}(0, 0.1)$. These new values are used as the inputs in Figure B.1 and the simulations are run for each station n_i

Bibliography

- Aab A., et al., 2017, *Science*, 357, 1266
- Andriopoulou M., Mavromichalaki H., Plainaki C., Belov A., Eroshenko E., 2011, *Sol Phys*, 269, 155
- Appourchaux T., 2004, *A&A*, 428, 1039
- Arunbabu K. P., et al., 2015, *Astronomy and Astrophysics*, 580, A41
- Aslam O. P. M., Badruddin 2012, *Solar Physics*, 279, 269
- Aslam O. P. M., Badruddin 2015, *Sol Phys*, 290, 2333
- Augusto C., Navia C., de Oliveira M. N., Fauth A., Nepomuceno A., 2016, *Publ Astron Soc Jpn Nihon Tenmon Gakkai*, 68
- BGS 2020, World Magnetic Model Calculator, http://www.geomag.bgs.ac.uk/data_service/models_compass/wmm_calc.html
- BIS 2015, Space weather preparedness strategy - GOV.UK. Vol. 2.1, <https://www.gov.uk/government/publications/space-weather-preparedness-strategy>
- Babcock H. W., 1953, *The Astrophysical Journal*, 118, 387
- Babcock H. W., 1961, *The Astrophysical Journal*, 133, 572
- Babcock H. W., Babcock H. D., 1955, *The Astrophysical Journal*, 121, 349
- Barnes a. W. T., et al., 2020, *ApJ*, 890, 68
- Bartels R. T., 2012, Technical report, The HiSPARC Experiment: An Analysis of the MPV and the Number of Events per Unit Time. University College Utrecht
- Bartels J., Heck N. H., Johnston H. F., 1939, *Terrestrial Magnetism and Atmospheric Electricity*, 44, 411
- Basu S., Chaplin W. J., 2017, Asteroseismic Data Analysis: Foundations and Techniques
- Beck J. G., 2000, *Solar Physics*, 191, 47
- Beggan C., Wild J., Gibbs M., 2018, *A&G*, 59, 4.36
- Belov A., 2000, *Space Science Reviews*, 93, 79

- Belov A. V., 2008, Proceedings of the International Astronomical Union, 4, 439
- Belov A. V., Eroshenko E. A., Oleneva V. A., Struminsky A. B., Yanke V. G., 2001, Advances in Space Research, 27, 625
- Belov A., Eroshenko E., Mavromichalaki H., Plainaki C., Yanke V., 2005, Annales Geophysicae, 23, 2281
- Belov A. V., Eroshenko E. A., Kryakunova O. N., Kurt V. G., Yanke V. G., 2010, Geomagn. Aeron., 50, 21
- Belov A., et al., 2014, Solar Physics, 289, 3949
- Belov A., Eroshenko E., Yanke V., Oleneva V., Abunin A., Abunina M., Papaioannou A., Mavromichalaki H., 2018, Sol Phys, 293, 68
- Berkova M. D., Belov A. V., Eroshenko E. A., Yanke V. G., 2011, Bull. Russ. Acad. Sci. Phys., 75, 820
- Bethe H., 1932, Zeitschrift fr Physik, 76, 293
- Blackmore E. W., Stukel M., Trinczek M., Slayman C., Wen S., Wong R., 2015, IEEE Transactions on Nuclear Science, 62, 2792
- Bose S., Nagaraju K., 2018, The Astrophysical Journal, 862, 35
- Braun I., Engler J., Hrandel J. R., Milke J., 2009, Advances in Space Research, 43, 480
- Brookes J. R., Isaak G. R., Raay H. B. v. d., 1976, Nature, 259, 92
- Brookes J. R., Isaak G. R., van der Raay H. B., 1978, Mon Not R Astron Soc, 185, 1
- Broomhall A.-M., 2017, Sol Phys, 292, 67
- Brown T. M., Christensen-Dalsgaard J., Dziembowski W. A., Goode P., Gough D. O., Morrow C. A., 1989, The Astrophysical Journal, 343, 526
- Bruno A., et al., 2018, The Astrophysical Journal, 862, 97
- CAEN 2011, User's Manual 00103/00:844-5.MUTx/07, Technical Information Manual: MOD. N844/N844P/N845 8-16 CHANNEL LOW THRESHOLD DISCRIMINATOR. CAEN
- Cabinet Office 2017, National Risk Register of Civil Emergencies - GOV.UK, 2017 edn. <https://www.gov.uk/government/publications/national-risk-register-of-civil-emergencies-2017-edition>
- Cane H. V., 2000, Space Science Reviews, 93, 55
- Canfield R. C., 2001, in Murdin P., ed., , The Encyclopedia of Astronomy and Astrophysics. IOP Publishing Ltd, doi:10.1888/0333750888/2023, <http://eaa.crcpress.com/0333750888/2023>
- Cannon P., Royal Academy of Engineering (Great Britain) 2013, Extreme space weather: impacts on engineered systems and infrastructure : summary report

- Carrington R. C., 1859, *Monthly Notices of the Royal Astronomical Society*, 20, 13
- Carrington R. C., 1863, Observations of the spots on the sun from November 9, 1853, to March 24, 1861, made at Redhill. London [etc.] Williams and Norgate, <http://archive.org/details/observationsofsp00carr>
- Casanovas J., 1997, 118, 3
- Casas R., Vaquero J. M., Vazquez M., 2006, *Solar Physics*, 234, 379
- Cecchini S., Sioli M., 2000, arXiv:hep-ex/0002052
- Chaplin W. J., et al., 1996, *Solar Physics*, 168, 1
- Chaplin W. J., Dumbill A. M., Elsworth Y., Isaak G. R., McLeod C. P., Miller B. A., New R., Pintr B., 2003, *Mon Not R Astron Soc*, 343, 813
- Chaplin W. J., Elsworth Y., Isaak G. R., Miller B. A., New R., Pintr B., 2005, *Mon Not R Astron Soc*, 359, 607
- Chaplin W. J., Elsworth Y., New R., Toutain T., 2008, *Monthly Notices of the Royal Astronomical Society*, 384, 1668
- Chaplin W. J., et al., 2019, *Monthly Notices of the Royal Astronomical Society: Letters*, 489, L86
- Chapman S. C., McIntosh S. W., Leamon R. J., Watkins N. W., 2020, *Geophysical Research Letters*, 47, e2020GL087795
- Charbonneau P., 2020, *Living Reviews in Solar Physics*, 17, 4
- Chilingarian A., 2003. CERN, Tsakhkadzor, Armenia, pp 63–79, <https://cds.cern.ch/record/604025/files/CERN-2005-007.pdf#page=73>
- Clark D. H., Stephenson F. R., 1978, *Quarterly Journal of the Royal Astronomical Society*, 19, 387
- Clem J. M., Dorman L. I., 2000, *Space Science Reviews*, 93, 335
- Clette F., Lefvre L., 2016, *Solar Physics*, 291, 2629
- Clette F., Cliver E. W., Lefvre L., Svalgaard L., Vaquero J. M., Leibacher J. W., 2016, *Solar Physics*, 291, 2479
- Corti C., Potgieter M. S., Bindi V., Consolandi C., Light C., Palermo M., Popkow A., 2019, *ApJ*, 871, 253
- Covington A. E., 1969, *Journal of the Royal Astronomical Society of Canada*, 63, 125
- Cramp J. L., 1996, PhD Thesis, University of Tasmania, Hobart, <https://core.ac.uk/download/pdf/33327719.pdf>
- Cramp, J. L. Duldig, M. L. Flckiger, E. O. Humble, J. E. Shea, M. A. Smart, D. F. 1997, *Journal of Geophysical Research: Space Physics*, 102, 24237

- Dacie S., Dmoulin P., Driell-Gesztesy L. v., Long D. M., Baker D., Janvier M., Yardley S. L., Prez-Surez D., 2016, *A&A*, 596, A69
- Danilova O. A., Demina I. M., Ptitsyna N. G., Tyasto M. I., 2019, *Geomagnetism and Aeronomy*, 59, 147
- Dasi-Espuig M., Solanki S. K., Krivova N. A., Cameron R., Peuela T., 2010, *Astronomy & Astrophysics*, 518, A7
- Davies G. R., Broomhall A. M., Chaplin W. J., Elsworth Y., Hale S. J., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 2025
- De Mendonça R. R. S., Raulin J. P., Echer E., Makhmutov V. S., Fernandez G., 2013, *J. Geophys. Res. Space Physics*, 118, 1403
- Desorgher L., 2005, Technical Report PLANETOCOSMICS-SUM, PLANETOCOSMICS Software User Manual, <http://cosray.unibe.ch/~laurent/planetocosmics/>. University of Bern
- Desorgher L., Flückiger E. O., Gurtner M., 2006. p. 2361, <http://adsabs.harvard.edu/abs/2006cosp...36.2361D>
- Dorman L. I., 1972, Kosm. Luchi, no. 13, pp. 5-66
- Dorman L. I., 2004a, in *Astrophysics and Space Science Library, Cosmic Rays in the Earth's Atmosphere and Underground*. Springer, Dordrecht, pp 289–330, doi:10.1007/978-1-4020-2113-8_5, https://link.springer.com/chapter/10.1007/978-1-4020-2113-8_5
- Dorman L. I., 2004b, in *Astrophysics and Space Science Library, Cosmic Rays in the Earth's Atmosphere and Underground*. Springer, Dordrecht, pp 375–386, doi:10.1007/978-1-4020-2113-8_7, https://link.springer.com/chapter/10.1007/978-1-4020-2113-8_7
- Dorman L. I., 2010, *Phys.-Usp.*, 53, 496
- Dubey A., Kumar S., Dubey S. K., 2016, *Journal of Physics: Conference Series*, 755, 012056
- Duldig M. L., et al., 1993, *Proceedings of the Astronomical Society of Australia*, 10, 211
- Dumbill A. M., 1999, PhD thesis, School of Physics and Space Research, University of Birmingham
- Dumbović M., Vršnak B., Čalogović J., Župan R., 2012, *Astronomy & Astrophysics*, 538, A28
- Dumbović M., Vršnak B., Čalogović J., 2016, *Sol Phys.*, 291, 285
- Dunai T. J., 2010, in *Cosmogenic Nuclides: Principles, Concepts and Applications in the Earth Surface Sciences*. Cambridge University Press, Cambridge, pp 1–24, doi:10.1017/CBO9780511804519.003
- Dyer C. S., Lei F., Clucas S. N., Smart D. F., Shea M. A., 2003, *Advances in Space Research*, 32, 81

ETEnterprises 2020, Technical Report 9125B, Data Sheet: 9125B Series, <http://et-enterprises.com/products/photomultipliers/product/p9125b-series>. ET Enterprises

Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., van der Raay H. B., Wheeler S. J., New R., 1995a. p. 392, <http://adsabs.harvard.edu/abs/1995ASPC...76..392E>

Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., New R., Wheeler S. J., 1995b, *Astronomy and Astrophysics Supplement Series*, 113, 379

Fan K.-z., Velthuis J., 2018, arXiv:1808.10645 [astro-ph]

Fesefeldt H., 1985, Technical Report PTHA-85/02, GHEISHA. The Simulation of Hadronic Showers: Physics and Applications. RWTH Aachen

Fokkema D. B. R. A., 2012, PhD thesis, University of Twente

Fokkema D. B. R. A., 2019, HiSPARC station software en hardware documentatie HiSPARC software 6.11 documentation, <https://docs.hisparc.nl/station-software/doc/index.html>

Fokkema D. B. R. A., de Laat A. P. L. S., Kooij T., 2012, SAPPHIRE, <https://github.com/HiSPARC/sapphire>

Forbush S. E., 1937, *Physical Review*, 51, 1108

Forbush S. E., 1958, *Journal of Geophysical Research*, 63, 651

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306

Gaisser T. K., Engel R., Resconi E., 2016, *Cosmic Rays and Particle Physics*, 2nd edn. Cambridge University Press, doi:10.1017/CBO9781139192194

Garca R. A., et al., 1999, *Astronomy and Astrophysics*, 346, 626

Gelman A., Rubin D. B., 1992, *Statistical Science*, 7, 457

Giacalone J., 2010, in Schrijver C. J., Siscoe G. L., eds, , *Heliophysics: Space Storms and Radiation: Causes and Effects*. Cambridge University Press, Cambridge, pp 233–262, doi:10.1017/CBO9781139194532.010

Gloeckler G., 2010, in Schrijver C. J., Siscoe G. L., eds, , *Heliophysics: Space Storms and Radiation: Causes and Effects*. Cambridge University Press, Cambridge, pp 233–262, doi:10.1017/CBO9781139194532.004

Gnevyshev M. N., 1938, *Pulkovo Obs. Circ*, 24, 37

Gnevyshev M. N., 1963, *Soviet Astronomy*, 7, 311

Gnevyshev M. N., 1967, *Solar Physics*, 1, 107

Gopalswamy N., 2010, pp 108–130

Gopalswamy N., Yashiro S., Michalek G., Stenborg G., Vourlidas A., Freeland S., Howard R., 2009, *Earth, Moon, and Planets*, 104, 295

Government H., 2020, Risk Register, National Risk Register: 2020 edition, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/952959/6.6920_CO_CCS_s_National_Risk_Register_2020_11-1-21-FINAL.pdf. HM Government

Groom D. E., Mokhov N. V., Striganov S. I., 2001, *Atomic Data and Nuclear Data Tables*, 78, 183

Grupen C., 2005, Astroparticle Physics. Springer-Verlag, Berlin Heidelberg, doi:10.1007/3-540-27670-X, <https://www.springer.com/gp/book/9783540253129>

Hagenaar H. J., 2001, *ApJ*, 555, 448

Hale G. E., Nicholson S. B., 1925, *The Astrophysical Journal*, 62, 270

Hale G. E., Ellerman F., Nicholson S. B., Joy A. H., 1919, *The Astrophysical Journal*, 49, 153

Hale S. J., Howe R., Chaplin W. J., Davies G. R., Elsworth Y., 2016, *Solar Physics*, 291, 1

Hanasoge S., Mandal K., 2019, *The Astrophysical Journal Letters*, 871, L32

Handberg R., Campante T. L., 2011, *Astronomy & Astrophysics*, 527, A56

Harvey K., 2001, in , The Encyclopedia of Astronomy and Astrophysics. IOP Publishing Ltd, doi:10.1888/0333750888/2275, <http://eaa.crcpress.com/0333750888/2275>

Harvey K. L., Martin S. F., 1973, *Solar Physics*, 32, 389

Harvey K. L., Zwaan C., 1993, *Solar Physics*, 148, 85

Hathaway D. H., 2015, *Living Rev. Sol. Phys.*, 12, 4

Hathaway D. H., Choudhary D. P., 2008, *Sol Phys*, 250, 269

Hathaway D. H., Upton L. A., 2017, Solar Cycle Science, <http://solarcyclescience.com/index.html>

Hathaway D. H., Upton L. A., 2021, *The Astrophysical Journal*, 908, 160

Heck D., Pierog T., 2017, Technical Report 7.6400, Extensive Air Shower Simulation with CORSIKA: A Users Guide

Herbst K., Kopp A., Heber B., 2013, *Annales Geophysicae*, 31, 1637

Homeier N., Wei L., 2013, Technical report, Solar Storm Risk to the North American Electric Grid, <https://www.lloyds.com/news-and-risk-insight/risk-reports/library/natural-environment/solar-storm>. Lloyd's

Howard R. F., 2001, in , The Encyclopedia of Astronomy and Astrophysics. IOP Publishing Ltd, doi:10.1888/0333750888/2297, <http://eaa.crcpress.com/0333750888/2297>

- Howe R., 2009, *Living Reviews in Solar Physics*, 6
- Howe R., Hill F., Komm R., Chaplin W. J., Elsworth Y., Davies G. R., Schou J., Thompson M. J., 2018, *ApJL*, 862, L5
- Howe R., et al., 2020, *Mon Not R Astron Soc Lett*, 493, L49
- Humble J. E., Duldig M. L., Smart D. F., Shea M. A., 2012, *Geophysical Research Letters*, 18, 737
- Inceoglu F., Knudsen M. F., Karoff C., Olsen J., 2014, *Sol Phys*, 289, 1387
- Janardhan P., Fujiki K., Ingale M., Bisoi S. K., Rout D., 2018, *Astronomy and Astrophysics*, 618, A148
- Kane R. P., 2014, *Solar Physics*, 289, 2727
- Kang J., Jang D. Y., Kim Y., Kang B. H., Kim Y.-K., Kim J., Park H., Yi Y., 2012, *Journal of the Korean Physical Society*, 61, 720
- Keller C. U., Schssler M., Vögler A., Zakharov V., 2004, *ApJ*, 607, L59
- Kiepenheuer K. O., ed. 1968, Structure and Development of Solar Active Regions. International Astronomical Union Symposia, Springer Netherlands, doi:10.1007/978-94-011-6815-1, <https://www.springer.com/gp/book/9789027701220>
- Knipp D. J., et al., 2016, *Space Weather*, 14, 2016SW001423
- Kotov V. A., 2008, *Astron. Rep.*, 52, 419
- Kotov V. A., 2012, *Bull.Crim. Astrophys. Observ.*, 108, 20
- Kudela K., Storini M., Hofer M. Y., Belov A., 2000, *Space Science Reviews*, 93, 153
- Kutsenko A. S., Abramenko V. I., Yurchyshyn V. B., 2017, *Sol Phys*, 292, 121
- Kuwabara T., et al., 2006a, *Space Weather*, 4, S08001
- Kuwabara T., Bieber J. W., Clem J., Evenson P., Pyle R., 2006b, *Space Weather*, 4, S10001
- LISIRD 2019, Solar Radio Flux at 10.7cm, Time Series, https://lasp.colorado.edu/lisird/data/penticton_radio_flux/
- Lanza A. F., Gizon L., Zaqrashvili T. V., Liang Z.-C., Rodenbeck K., 2019, *A&A*, 623, A50
- LeCroy 1996, LeCroy Research Systems Model 365AL, Model 465, and Technical Data: Model 622 Logic Units, <https://teledynelecroy.com/lrs/dsheets/365al.htm>
- Leamon R. J., McIntosh S. W., Chapman S. C., Watkins N. W., 2020, *Solar Physics*, 295, 36
- Lee 1858, *Monthly Notices of the Royal Astronomical Society*, 19, 1

- Lei F., Clucas S., Dyer C., Truscott P., 2004, IEEE Transactions on Nuclear Science, 51, 3442
- Leighton R. B., 1964, The Astrophysical Journal, 140, 1547
- Li K. J., Yun H. S., Gu X. M., 2001, AJ, 122, 2115
- Liang Z.-C., Gizon L., Birch A. C., Duvall T. L., 2019, A&A, 626, A3
- Lingri D., Mavromichalaki H., Belov A., Eroshenko E., Yanke V., Abunin A., Abunina M., 2016, arXiv:1612.08900 [astro-ph]
- Lista L., 2016, Statistical Methods for Data Analysis in Particle Physics. Lecture Notes in Physics, Springer International Publishing, doi:10.1007/978-3-319-20176-4, <https://www.springer.com/gp/book/9783319201764>
- Lockwood J. A., 1971, Space Science Reviews, 12, 658
- Lockwood M., Stamper R., Wild M. N., 1999, Nature, 399, 437
- Löptien B., Gizon L., Birch A. C., Schou J., Proxauf B., Duvall T. L., Bogart R. S., Christensen U. R., 2018, Nature Astronomy, 2, 568
- Lovell J. L., Duldig M. L., Humble J. E., 1998, Journal of Geophysical Research: Space Physics, 103, 23733
- Luminet J.-P., 2017, International Review of Science
- Lund M. N., Chaplin W. J., Hale S. J., Davies G. R., Elsworth Y. P., Howe R., 2017, Mon Not R Astron Soc, 472, 3256
- Martin S. F., 1988, Solar Physics, 117, 243
- Martin S. F., Harvey K. L., 1979, Solar Physics, 64, 93
- Maunder E. W., 1903, The Observatory, 26, 329
- Maunder E. W., 1904, Mon Not R Astron Soc, 64, 747
- Mavromichalaki H., Paouris E., Karalidi T., 2007, Sol Phys, 245, 369
- McCracken K. G., Rao U. R., Shea M. A., 1962, Technical Report NYO-2670, THE TRAJECTORIES OF COSMIC RAYS IN A HIGH DEGREE SIMULATION OF THE GEOMAGNETIC FIELD. Technical Report No. 77, <https://www.osti.gov/biblio/4784046>. Massachusetts Inst. of Tech., Cambridge. Lab. for Nuclear Science
- McCracken K. G., Moraal H., Shea M. A., 2012, ApJ, 761, 101
- McIntosh S. W., et al., 2014, ApJ, 792, 12
- McIntosh S. W., Leamon R. J., Egeland R., Dikpati M., Fan Y., Rempel M., 2019, Sol Phys, 294, 88
- McIntosh S. W., Chapman S., Leamon R. J., Egeland R., Watkins N. W., 2020, Solar Physics, 295, 163

- Mendonça R. R. S. d., et al., 2016, *ApJ*, 830, 88
- MetOffice 2013, Technical Report 13/0402, Space Weather, https://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/public-sector/emergencies/space-weather/space_weather_datasheet_final.pdf
- Mishev A., Usoskin I., 2020, *Journal of Space Weather and Space Climate*, 10, 17
- Mishev A. L., Usoskin I. G., Raukunen O., Paassilta M., Valtonen E., Kocharov L. G., Vainio R., 2018, *Solar Physics*, 293
- Mishra R. K., Mishra R. A., 2007, *Pramana*, 68, 407
- Mishra R. K., Mishra R. A., 2008, *Rom. Journ. Phys.*, 53, 925
- Mishra V. K., Mishra A. P., 2016, *Indian J Phys*, 90, 1333
- Montanus H., 2017, PhD thesis, University of Amsterdam, https://www.nikhef.nl/pub/services/biblio/theses_pdf/
- Moraal H., Caballero-Lopez R. A., 2014, *The Astrophysical Journal*, 790, 154
- NMDB 2018, NMDB Event Search Tool (NEST), <http://www.nmdb.eu/nest/>
- NOAA 2018, NOAA Space Weather Predictions Center, <https://www.swpc.noaa.gov/>
- Nature 1871, Observations Upon Magnetic Storms in Higher Latitudes, doi:10.1038/004441a0, <https://www.nature.com/articles/004441a0>
- Nelson W. R., Hirayama H., Rogers D. W. O., 1985, Technical Report SLAC-265, EGS4 code system, <https://www.osti.gov/biblio/6137659-egs4-code-system>. Stanford Linear Accelerator Center, Menlo Park, CA (USA)
- Norton A. A., Gallagher J. C., 2009, *Solar Physics*, 261, 193
- Ostapchenko S., 2006, *Nuclear Physics B - Proceedings Supplements*, 151, 143
- Owens M. J., Forsyth R. J., 2013, *Living Rev. Sol. Phys.*, 10, 5
- Pacini A. A., Usoskin I. G., 2015, *Sol Phys*, 290, 943
- Paouris E., Mavromichalaki H., Belov A., Eroshenko E., Gushchina R., 2015, *J. Phys.: Conf. Ser.*, 632, 012074
- Parker E. N., 1958, *The Astrophysical Journal*, 128, 664
- Parker E. N., 1964, *Planetary and Space Science*, 12, 735
- Parker E. N., 1965, *Planetary and Space Science*, 13, 9
- Parker G. D., Jokipii J. R., 1976, *Geophysical Research Letters*, 3, 561
- Particle Data Group et al., 2020, *Progress of Theoretical and Experimental Physics*, 2020
- Paschalis P., Mavromichalaki H., Yanke V., Belov A., Eroshenko E., Gerontidou M., Koutroumpi I., 2013, *New Astronomy*, 19, 10

- Pereira E. G., Vasconcellos C. A., Hadjimichef D., 2021, *Astronomische Nachrichten*
- Pesnell W. D., 2020, *Journal of Space Weather and Space Climate*, 10, 60
- Pesnell W. D., Schatten K. H., 2018, *Sol Phys*, 293, 112
- Pillet V. M., Lites B. W., Skumanich A., 1997, *ApJ*, 474, 810
- Plachinda S., Pankov N., Baklanova D., 2011, *Astronomische Nachrichten*, 332, 918
- Plainaki C., Mavromichalaki H., Belov A., Eroshenko E., Yanke V., 2009, *Advances in Space Research*, 43, 518
- Poluijanov S., Usoskin I. G., 2017, in Proceedings of Science. Bexco, Busan, Korea, <https://pos.sissa.it/301/#session-3145>
- Poluijanov S. V., Usoskin I. G., Mishev A. L., Shea M. A., Smart D. F., 2017, *Sol Phys*, 292, 176
- Riley P., 2012, *Space Weather*, 10
- Rockenbach M., et al., 2014, *Space Sci Rev*, 182, 1
- Ross E., Chaplin W. J., 2019, *Sol Phys*, 294, 8
- Ross E., Chaplin W. J., Hale S. J., Howe R., Elsworth Y. P., Davies G. R., Nielsen M. B., 2021, *Monthly Notices of the Royal Astronomical Society*
- Rossby C.-G., Collaborators 1939, *J Mar Res*, 2, 38
- Russell C. T., et al., 2013, *ApJ*, 770, 38
- SILSO WDC 2020, International Sunspot Number Monthly Bulletin and online catalogue
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Comput. Sci.*, 2, e55
- Sanchez S., Fournier A., Pinheiro K. J., Aubert J., Sanchez S., Fournier A., Pinheiro K. J., Aubert J., 2014, *Anais da Academia Brasileira de Ciencias*, 86, 11
- Savić M., Maletić D., Joković D., Veselinović N., Banjanac R., Udovičić V., Dragić A., 2015, *J. Phys.: Conf. Ser.*, 632, 012059
- Scherrer P. H., Wilcox J. M., Howard R., 1972, *Sol Phys*, 22, 418
- Scherrer P. H., Wilcox J. M., Kotov V., Severnyj A. B., Severny A. B., Howard R., 1977a, *Solar Physics*, 52, 3
- Scherrer P. H., Wilcox J. M., Svalgaard L., Duvall Jr. T. L., Dittmer P. H., Gustafson E. K., 1977b, *Solar Physics*, 54, 353
- Schrijver C. J., Harvey K. L., 1994, *Solar Physics*, 150, 1
- Schrijver C. J., Siscoe G. L., eds, 2010, *Heliophysics: Space Storms and Radiation: Causes and Effects*. Cambridge University Press, Cambridge, doi:10.1017/CBO9781139194532, <https://www.cambridge.org/core/books/heliophysics-space-storms-and-radiation-causes-and-effects/608857F4EEE1C0BF61F26ACD19BA59DE>

- Schrijver C. J., Zwaan C., 2008, Solar and Stellar Magnetic Activity. Cambridge University Press
- Schwabe H., 1844, *Astronomische Nachrichten*, 21, 233
- Schwenn R., 2006, *Living Reviews in Solar Physics*, 3, 2
- Severny A. B., 1971, *Quarterly Journal of the Royal Astronomical Society*, 12, 363
- Shea M. A., 1990, 12, 196
- Shea M. A., Smart D. F., 1982, *Space Science Reviews*, 32, 251
- Shea M. A., Smart D. F., McCracken K. G., 1965, *J. Geophys. Res.*, 70, 4117
- Sheeley N. R., 2005, *Living Reviews in Solar Physics*, 2, 5
- Simpson J. A., 1948, *Phys. Rev.*, 73, 1389
- Simpson J. A., Fonger W., Treiman S. B., 1953, *Phys. Rev.*, 90, 934
- Singh M., Singh Y. P., Badruddin 2008, *Journal of Atmospheric and Solar-Terrestrial Physics*, 70, 169
- Snodgrass H. B., 1983, *The Astrophysical Journal*, 270, 288
- Solanki S. K., 2003, *The Astronomy and Astrophysics Review*, 11, 153
- Solanki S. K., Schssler M., Fligge M., 2000, *Nature*, 408, 445
- Solanki S. K., Schssler M., Fligge M., 2002, *A&A*, 383, 706
- Solovev A. A., Parfinenko L. D., Efremov V. I., Kirichek E. A., Korolkova O. A., 2019, *Astrophysics and Space Science*, 364, 222
- Spruit H. C., 1976, *Solar Physics*, 50, 269
- Stancik A. L., Brauns E. B., 2008, *Vibrational Spectroscopy*, 47, 66
- Stoker P. H., Dorman L. I., Clem J. M., 2000, *Space Science Reviews*, 93, 361
- Strauss R. D., Ogunjobi O., Moraal H., McCracken K. G., Caballero-Lopez R. A., 2017, *Solar Physics*, 292, 51
- Strohmayer T., Mahmoodifar S., 2014, A Neutron Star Oscillation Mode During a Superburst, <https://slideplayer.com/slide/10313230/>
- Svalgaard L., Wilcox J. M., Scherrer P. H., Howard R., 1975, *Sol Phys*, 45, 83
- Tapping K. F., 2013, *Space Weather*, 11, 394
- Tapping K. F., Charrois D. P., 1994, *Solar Physics*, 150, 305
- Tapping K. F., DeTracey B., 1990, *Solar Physics*, 127, 321
- Thébault E., et al., 2015, *Earth, Planets and Space*, 67, 79

- Thomas S. R., Owens M. J., Lockwood M., 2014a, *Sol Phys*, 289, 407
- Thomas S. R., Owens M. J., Lockwood M., Scott C. J., 2014b, *Solar Physics*; Dordrecht, 289, 2653
- Thomas S., Owens M., Lockwood M., Owen C., 2017, *Ann. Geophys.*, 35, 825
- Thomas A. E. L., et al., 2019, *Mon Not R Astron Soc*, 485, 3857
- Timashkov D. A., et al., 2008, *Astroparticle Physics*, 30, 117
- Tomassetti N., Orcinha M., Baro F., Bertucci B., 2017, *ApJL*, 849, L32
- Tsyganenko N. A., 1989, *Planetary and Space Science*, 37, 5
- Tsyganenko N. A., 2013, *Annales Geophysicae*, 31, 1745
- Upton L. A., Hathaway D. H., 2018, *Geophysical Research Letters*
- Usoskin I. G., Kananen H., Mursula K., Tanskanen P., Kovaltsov G. A., 1998, *Journal of Geophysical Research: Space Physics*, 103, 9567
- Usoskin I. G., Braun I., Gladysheva O. G., Hörandel J. R., Jmsn T., Kovaltsov G. A., Starodubtsev S. A., 2008, *J. Geophys. Res.*, 113, A07102
- Usoskin I., Ibragimov A., Shea M. A., Smart D., 2016, in Proceedings of The 34th International Cosmic Ray Conference PoS(ICRC2015). Sissa Medialab, The Hague, The Netherlands, p. 054, doi:10.22323/1.236.0054, <https://pos.sissa.it/236/054>
- Van Allen J. A., 2000, *Geophysical Research Letters*, 27, 2453
- Vieira L. E. A., Solanki S. K., 2010, *A&A*, 509, A100
- Viljanen A., 2011, *Space Weather*, 9, 07007
- Waldmeier M., 1955, *Quarterly Journal of the Royal Meteorological Society*, 82, 118
- Wawrzynczak A., Alania M. V., 2010, *Advances in Space Research*, 45, 622
- Webb D. F., Howard R. A., 1994, *Journal of Geophysical Research: Space Physics*, 99, 4201
- Wilson P. R., 1994, Solar and Stellar Activity Cycles. Cambridge Astrophysics, Cambridge University Press, Cambridge, doi:10.1017/CBO9780511564833, <https://www.cambridge.org/core/books/solar-and-stellar-activity-cycles/49EFD2ED430403044DEB7B7177F77DAF>
- Wolf R., 1856, *Astronomische Mitteilungen der Eidgenössischen Sternwarte Zurich*, 1, 3
- Wolf R., 1859, *Monthly Notices of the Royal Astronomical Society*, 19, 85
- Wolf R., 1861, *Monthly Notices of the Royal Astronomical Society*, 21, 77
- Wu C. J., Usoskin I. G., Krivova N., Kovaltsov G. A., Baroni M., Bard E., Solanki S. K., 2018, *A&A*, 615, A93

- Xiang N. B., Qu Z. N., 2016, AJ, 151, 76
- Xie J. L., Shi X. J., Xu J. C., 2017, The Astronomical Journal, 153, 171
- Yashiro S., Gopalswamy N., Michalek G., Cyr O. C. S., Plunkett S. P., Rich N. B., Howard R. A., 2004, Journal of Geophysical Research: Space Physics, 109
- Ziegler J. F., 1999, Journal of Applied Physics, 85, 1249
- Zwaan C., 1981, NASA Special Publication, 450
- eFoam 2017, SF38 High Foam Datasheet, <https://www.efoam.co.uk/docs/SF38.pdf>
- van Dam K., van Eijk B., Steijger J. J. M., 2020a, arXiv:2009.05308 [astro-ph]
- van Dam K., van Eijk B., Fokkema D. B. R. A., van Holten J. W., de Laat A. P. L. S., Schultheiss N. G., Steijger J. J. M., Verkooijen J. C., 2020b, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 959, 163577
- van Driel-Gesztelyi L., Green L. M., 2015, Living Rev. Sol. Phys., 12, 1

