

Thesis Title

by

E. 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
Month 20XX

Contents

List of Figures	ix
List of Tables	xi
List of Abbreviations	xiii
1 HiSPARC as a Space Weather Detector	1
1.1 Introduction	1
1.2 Aims	7
1.3 HiSPARC Data	8
1.4 Station Properties	11
1.4.1 Cut-Off Rigidity	11
1.4.2 Asymptotic Viewing Directions	14
1.5 HiSPARC Observations	15
1.5.1 Observations of Ground Level Enhancements	15
1.5.2 Observations of Forbush Decreases	18
1.6 Atmospheric Corrections of HiSPARC Data	21
1.6.1 Motivation	21
1.6.2 Barometric Correction	23
1.6.3 Temperature Correction	26
1.7 HiSPARC Observations After Atmospheric Corrections	29
1.7.1 Observations of Ground Level Enhancements	29
1.7.2 Observations of Forbush Decreases	30
1.8 Air Shower Simulations	33
1.8.1 Motivation	33
1.8.2 Air Shower Footprints	33
1.8.3 Muon Flux	35
1.8.4 Muon Flux From MAIRE	38
1.9 Conclusion	43
2 HiSPARC Station 14008	47
2.1 Introduction	47
2.2 Aims	48
2.3 HiSPARC Station 14008 Set-up	48
2.3.1 Configuration	48
2.3.2 Calibration	53
2.3.3 Monitoring Temperature	55

2.3.4	Data Acquisition	55
2.3.5	Monitoring Pressure	58
2.4	Methodology	59
2.4.1	Atmospheric Corrections	59
2.4.2	Observations	60
2.5	Atmospheric Corrections	63
2.5.1	Temperature Correction	63
2.5.2	Barometric Correction	67
2.6	Results	69
2.6.1	Observations	69
2.6.2	Comparison with Neutron Monitors	74
2.6.3	Single Station Space Weather Uses	75
2.6.4	Multiple Station Space Weather Uses	81
2.7	Conclusions	87
A	Simulations of Cosmic Ray Air Showers using CORSIKA	91
B	Simulations of the Ground Level Enhancements for the Analysis of HiSPARC Station 14008	95
B.1	Model	95
B.2	Configuration of the Simulations	97
Bibliography		99

List of Figures

1.1	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.	6
1.2	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.	6
1.3	The geographic location of each HiSPARC station considered in this work. Each green circle denotes the location of a detector station.	10
1.4	The availability of data for each HiSPARC station considered, for each of the space weather epochs listed in Table 1.1. 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.	10
1.5	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.	13
1.6	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.	15
1.7	HiSPARC data for stations 501 and 3001 around the epoch of GLE 70 on 13/12/2006. The plot shows the minute-averaged and 5-minute-averaged 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.	16

1.8	HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged and 5-minute-averaged 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.	16
1.9	HiSPARC data for 4 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the 1- and 5-minute 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.	17
1.10	HiSPARC data for stations 501 and 8001 around the epoch of the FDs in March 2012. The plot shows the minute-averaged and hourly-averaged 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.	19
1.11	HiSPARC data for stations 501 and 8001 around the epoch of the FD in July 2012. The plot shows the minute-averaged and hourly-averaged 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.	19
1.12	HiSPARC data for four stations around the epoch of the FD in December 2014. The plot shows the minute-averaged and hourly-averaged 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.	20
1.13	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- and hourly-averaged 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.	22
1.14	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.	23
1.15	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.	25

1.16	A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the pressure correction (blue line).	25
1.17	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.	26
1.18	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.	28
1.19	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.	28
1.20	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 and 5-minute-averaged 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.	29
1.21	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 1- and 5-minute 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.	30
1.22	Atmospheric-corrected HiSPARC data for station 8001 (Eindhoven) around the epoch of the FD in July 2012. The plot shows the minute- and hourly-averaged 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.	31
1.23	Atmospheric-corrected HiSPARC data for station 501 (Nikhef) around the epoch of the FD in December 2014. The plot shows the minute- and hourly-averaged 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.	31

1.24 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- and hourly-averaged 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.	32
1.25 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	34
1.26 Mean number of muons produced at ground level by the PCR for (a) proton-initiated air showers and (b) α -particle-initiated air showers, for various PCR energies.	36
1.27 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°	37
1.28 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.	39
2.1 Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box.	49
2.2 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.	50
2.3 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 a HV signal. The PMTs from the detectors 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.	51

2.4	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)	54
2.5	(a) Schematic diagram of the DS18B20 temperature sensor circuit, whereby the voltage, ground, and GPIO interfaces connect directly into 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.	56
2.6	Schematic diagram of the HiSPARC station 14008 data acquisition interfaces. Black arrows depict a 5 V signal; green arrows show ground interfaces; blue arrows depict the 1-wire signal output from the temperature sensor	56
2.7	The relationship between the full coincidences data and the temperature within the roof box over a single day. (a) Shows the comparison between coincidences data (blue) and temperature data (orange); (b) shows the correlation between the coincidences counts and temperature, and the fitted line to calculate the correction coefficient.	63
2.8	The relationship between the full singles data (blue and orange lines) and the temperature within the roof box (black line).	64
2.9	The relationship between the singles data and the temperature within the roof box over a single day. (a) Shows the comparison between singles data (blue) and temperature data (orange); (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the correction coefficient.	65
2.10	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.	66
2.11	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.	67
2.12	The relationship between the smoothed, full coincidences data and the atmospheric pressure. (a) Shows the comparison between coincidences data (blue) and pressure data (orange), both with a 12-hour box-bar smoothing applied, to highlight the relationship; (b) shows the correlation between the coincidences counts and pressure, (blue points) and the fitted line (orange) to calculate the correction coefficient.	68
2.13	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.	68

2.14 Time series of coincidences data, corrected for atmospheric pressure. The blue line shows the corrected data displaying the diurnal variation with a peaks at around midday. The orange line shows the corrected coincidences de-trended using a 12-hour smoothing window, to remove the diurnal effect.	70
2.15 (a) Comparison of the reduced coincidences stored locally (orange) and as events data in the HiSPARC server (blue). (b) Distribution of the ratio of the full and reduced coincidence rates.	71
2.16 Time series of spurious coincidences data measured over five days. . .	72
2.17 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.	73
2.18 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, showing the similarities between the two data sets.	74
2.19 Single realisation of a simulation with a 10% Ground Level Enhancement (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 statistical significance. The dashed line shows the $p = 10\%$ threshold.	76
2.20 Summary plots showing the number of statistically significant 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.	79

2.21	Summary plots showing the number of statistically significant 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 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.	83
2.22	Cross-correlation analyses between 2, 5, and 10 stations for a 2% GLE magnitude with varying 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.	85
2.23	Cross-correlation analyses between 2, 5, and 10 stations for a 1% GLE magnitude with varying 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.	86
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.	96
B.2	Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.	97

List of Tables

1.1	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 at NMDB (2018) (for GLEs) and Lingri et al. (2016) (for FDs).	5
1.2	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.	13
1.3	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.	38
1.4	Observed and predicted increases in the CR count rates at the Kiel NM station, with rows ordered by the observed NM 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.	40
1.5	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 1.4. The rows ordered by the observed NM magnitude from Table 1.4.	42
2.1	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.	53
2.2	Variables stored in the coincidences files of the HiSPARC 14008 instrument.	57
2.3	Variables stored in the temperature files of the HiSPARC 14008 instrument.	57

2.4	Average number of statistically significant measurements in the artificial data with no injected GLEs. The values are for each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for the raw data and averaging over 1-minute and 5-minutes.	77
2.5	Average number of statistically significant measurements in the mean of the artificial data for multiple stations, with no injected GLEs. The values are for each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for tests with 1, 2, 5, and 10 stations.	82
A.1	Details of the proton-initiated air showers simulated using CORSIKA.	93
A.2	Details of the α -particle-initiated air showers simulated using CORSIKA.	94

List of Abbreviations

AMS-02 Alpha Magnetic Spectrometer.

AS Air Shower.

AVD Asymptotic Viewing Direction.

CCF Cross-Correlation Function.

CEDA Centre for Environmental Data Analysis.

CIR Corotating Interaction Region.

CME Coronal Mass Ejection.

CORSIKA Cosmic Ray Simulations for Kascade.

CR Cosmic Ray.

EAS Extensive Air Shower.

ESD Event Summary Data.

FD Forbush Decrease.

FE Forbush Effect.

FEID Forbush Effects and Interplanetary-disturbances Database.

FWHM Full Width at Half Maximum.

GCR Galactic Cosmic Ray.

GLE Ground Level Enhancement.

GNMN Global Neutron Monitor Network.

GPS Global Positioning System.

HiSPARC High School Project on Astrophysics and Research with Cosmics.

HMC Hamiltonian Monte Carlo.

- HV** High Voltage.
- ICME** Interplanetary Coronal Mass Ejection.
- IGRF** International Geomagnetic Reference Field.
- IMF** Interplanetary Magnetic Field.
- ISS** International Space Station.
- MAIRE** Model for Atmospheric Ionising Radiation Effects.
- MD** Muon Detector.
- MIDAS** Met Office Integrated Data Archive System.
- NERC** Natural Environment Research Council.
- NIM** Nuclear Instrumentation Module.
- NM** Neutron Monitor.
- NMDB** Neutron Monitor Data Base.
- NUTS** No U-Turn Sampler.
- PCR** Primary Cosmic Ray.
- PMT** Photo Multiplier Tube.
- SAPPHiRE** Simulation and Analysis Program Package for HiSPARC Research and Education.
- SCR** Solar Cosmic Ray.
- SEP** Solar Energetic Particle.
- STFC** Science and Technology Facilities Council.
- TTL** Transistor-Transistor Logic.
- UHECR** Ultra-High-Energy Cosmic Ray.

1 HiSPARC as a Space Weather Detector

1.1 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 environments 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; Poluianov & 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>

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. 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 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 by these two detectors and the signal at 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) claimed a 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, and highlights why NMs are routinely used.

The High School Project on Astrophysics and Research with Cosmics (HiSPARC)

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

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). 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., 2020). 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., 2020), but depending on the electron energies, they are also observable by HiSPARC.

The HiSPARC detectors have, individually, a high muon-detection efficiency close to 100% (van Dam et al., 2020), 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. This is due to the heritage of the HiSPARC network in observing Ultra-High-Energy Cosmic Rays (UHECRs). In this work, we provide a feasibility study to investigate whether the existing HiSPARC network is capable of observing space weather events.

Few space weather events have been observed over the lifetime of the HiSPARC network; however, Table 1.1 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 NMs: 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 1.1: 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 at [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

The specific events in Table 1.1 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.

For reference, and later comparison with the HiSPARC results, in Figure 1.1 we show the GLEs, as observed by the Oulu NM station, in Finland, using data taken from the NMDB. We see in Figure 1.1 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 1.2.

Each FD in Figure 1.2 produces a moderate decrease in the NM count rate. The relative decreases in the the CR counts during the FDs were generally around $\sim 5\%$, and are easily observable by-eye in the data.

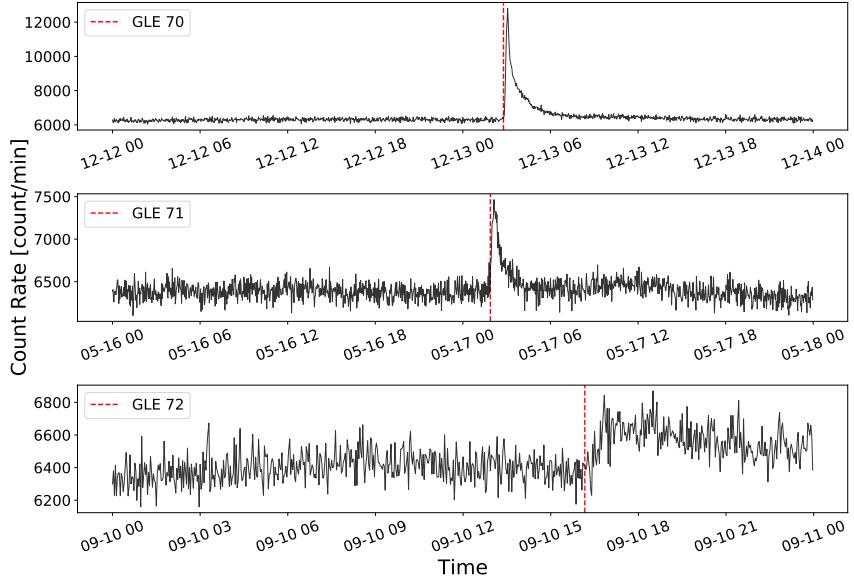


Figure 1.1: GLEs observed by the NM stations based at Oulu. Top panel: GLE 70; middle panel: GLE 71, bottom panel: GLE 72. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-red lines show the epochs of each GLE onset. The units of time on the x-axis are, MM-DD HH.

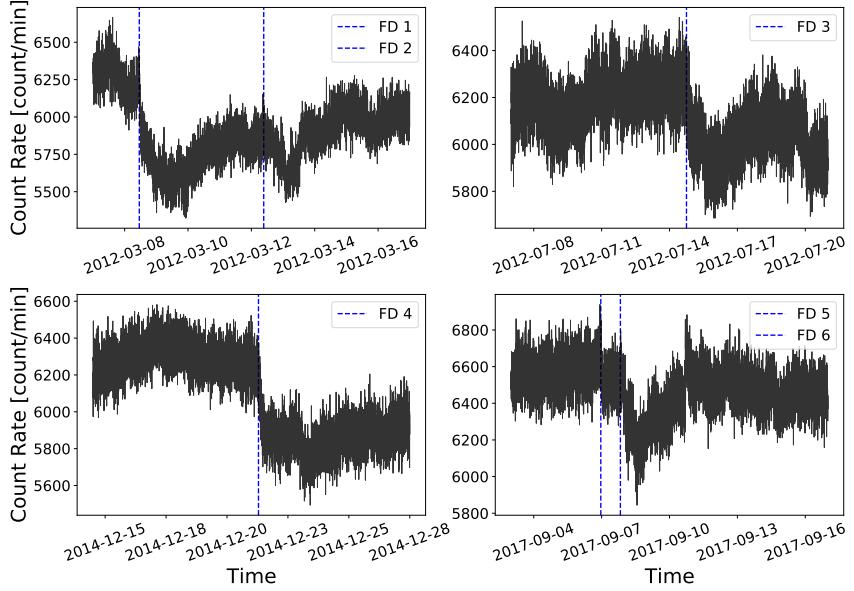


Figure 1.2: FDs observed by the NM stations based at Oulu. Top left panel: FDs during March 2012; top right panel: FD during July 2012; bottom left panel: FD during December 2014; bottom right panel: FD during September 2017. The solid-black line shows the 2-minute-averaged, pressure corrected data and the vertical, dashed-blue lines show the epochs of each FD onset. The units of time on the x-axis are, YYYY-MM-DD.

1.2 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 weather activity to search for the associated signatures detailed in Table 1.1. 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 forecasts 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.

1.3 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 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

³<https://data.hisparc.nl/>

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., 2020](#)) 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 1.2, a mixture of 2-detector and 4-detector stations. Approximately 110 of the \sim 140 stations record singles rates, which have only been available since 2016, and only 29 stations acquire weather 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 conditions. The stations are shown on a map in Fig. 1.3.

Figure 1.4 shows the availability of CR data at 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



Figure 1.3: The geographic location of each HiSPARC station considered in this work. Each green circle denotes the location of a detector station.

matures, so does the number of stations with available data, and the type of data available (both events and singles rates).

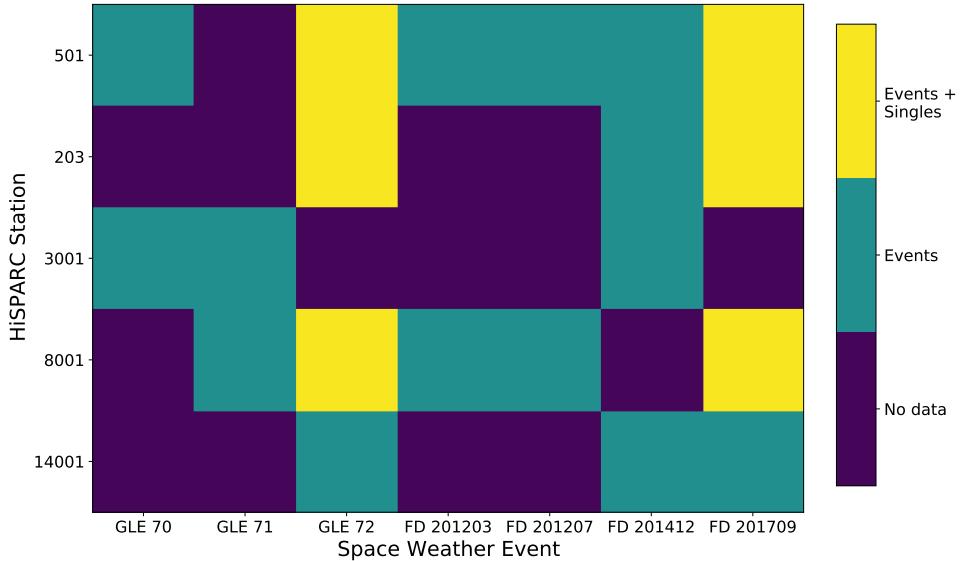


Figure 1.4: The availability of data for each HiSPARC station considered, for each of the space weather epochs listed in Table 1.1. The purple grids denote no available data, teal denotes that only the events data are available, and yellow denotes that both the events and singles data are available.

1.4 Station Properties

1.4.1 Cut-Off Rigidity

To understand the PCR spectrum that ground-based CR detectors may observe, PCR transport simulations are typically performed. 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 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 and 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 Størmers transport 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:

$$R_C = R_U - \sum_{i=R_L}^{R_U} \Delta R_i \quad (1.1)$$

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 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 1.2 which show that there is little variation in R_C between the HiSPARC stations and that they observe protons with rigidities in excess of ~ 3 GV.

Table 1.2: 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

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 1.5, and demonstrates that there is no strong dependence of the azimuth direction or zenith (up to 45°).

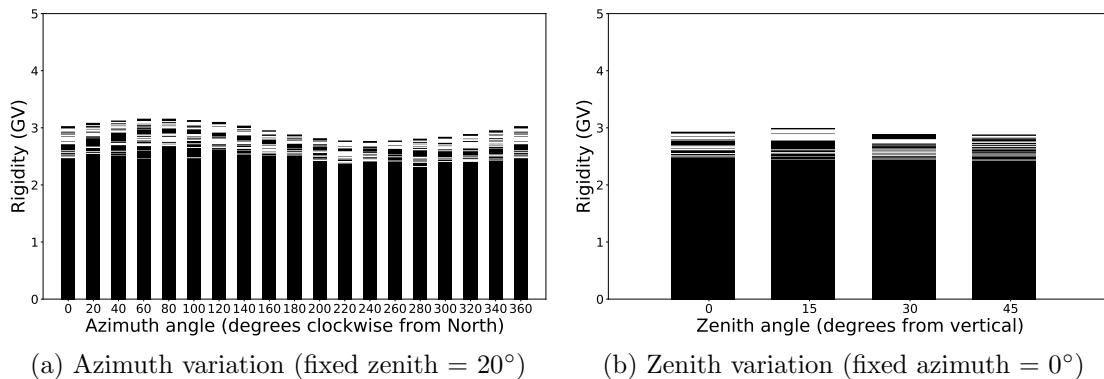


Figure 1.5: Azimuthal and zenith angle variations in the allowed and forbidden rigidity trajectories for HiSPARC station 501 from simulations in steps of rigidity, $\Delta R = 0.01$ GV. The forbidden trajectories are in black; allowed trajectories are in white.

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 Ray

(SCR), but may not be as susceptible as the higher latitude NMs where the effects of GLEs are highly observable.

1.4.2 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 ([Danilova et al., 2019](#)). This allowed us to understand the 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, there stations with a lower rigidity cut-off may observe CRs from a range of directions.

It can be seen from Figure 1.6 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, and the 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 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 $R_C \sim 2.5$ GV are shifted East by $\sim 120^\circ$ longitude. This demonstrates that the observable, lower-energy PCRs

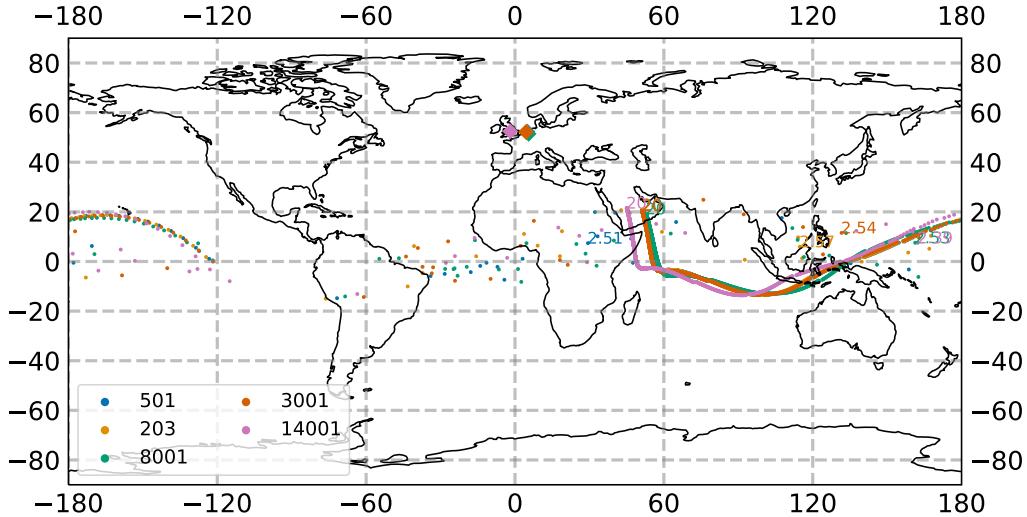


Figure 1.6: The vertical asymptotic viewing directions of 5 HiSPARC stations. The rigidity range of the simulations were from $1.0 \text{ GV} < R < 20.0 \text{ GV}$, and the results are plotted in geographic coordinates on January 20th 2005. The diamonds correspond to the HS ground location and the circles correspond to the AVD for a specific rigidity value.

are deflected significantly by the Earth's magnetosphere; we 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 significant impacts on the ability of our detectors to observe transient solar eruptive events which may be highly anisotropic and have a SEP spectrum with energies $< 10^9 \text{ eV}$. For this reason, many ground-based CR stations are spread across Earth's surface, to maximise the observation coverage.

1.5 HiSPARC Observations

1.5.1 Observations of Ground Level Enhancements

The search for evidence of GLEs within the HiSPARC data was conducted for the events listed in Table 1.1, as they are the only GLEs that span the operational epoch of the HiSPARC network. Figure 1.7, 1.8, and 1.9 show the HiSPARC observations around GLE 70, 71, and 72, respectively. As highlighted by Fig. 1.4, 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.

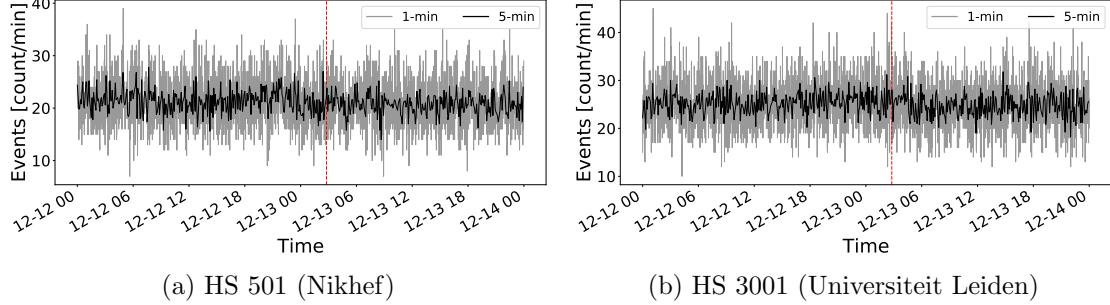


Figure 1.7: HiSPARC data for stations 501 and 3001 around the epoch of GLE 70 on 13/12/2006. The plot shows the minute-averaged and 5-minute-averaged trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

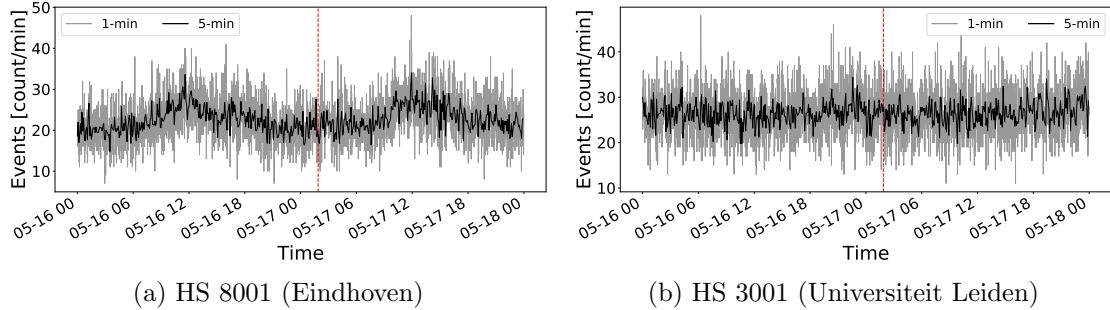


Figure 1.8: HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged and 5-minute-averaged trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

We can see from Figures 1.7, 1.8, and 1.9 there are no clear and obvious signs of the GLE signals in the HiSPARC observations, as was clear for those given in Fig. 1.1 for the Oulu NM station, in Finland. This is the case for both the events data and the singles data.

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. 1.9) for stations 501, 203, and 8001. We believe these excursions are the effect

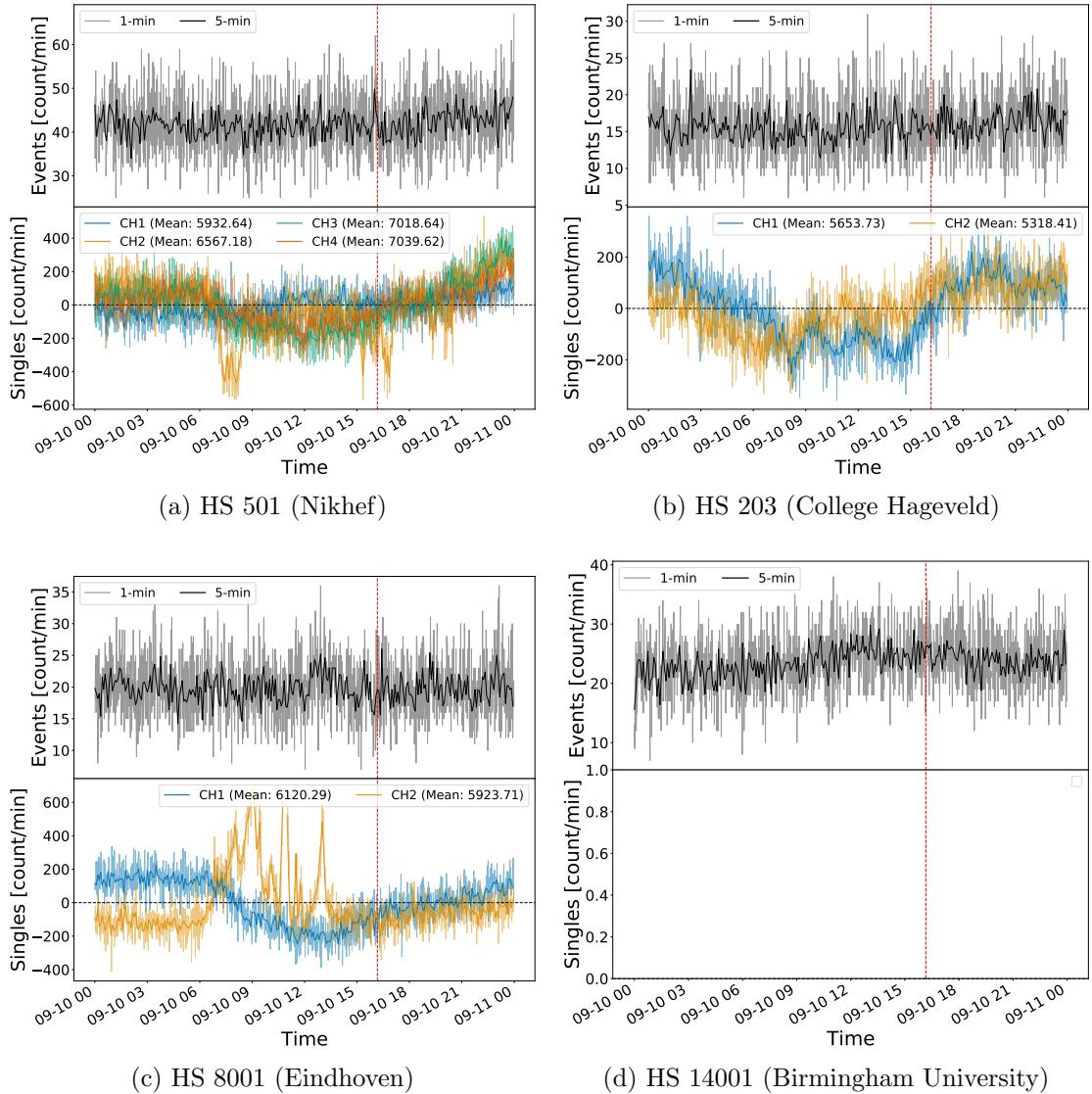


Figure 1.9: HiSPARC data for 4 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the 1- and 5-minute averaged trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, for each individual 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.

of atmospheric pressure and temperature on the muon count rates; in Section 1.6 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 1.7. We also see the existence of a diurnal signal in Figure 1.8a, 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 dif-

ferent 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 $\sim 50 - 100\%$, 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 NMAs. 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. Therefore, the rigidity cut-off of the HiSPARC stations is not necessarily the limiting factor. It is also possible that the PCRs are insufficient to produce EASs of muons, as 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 typically GLEs are observed by NMAs. 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 1.8 we used air shower simulations to investigate the CR spectrum at quiet times and during GLEs.

1.5.2 Observations of Forbush Decreases

The search for evidence of FDs within the HiSPARC data was conducted for the FDs highlighted in Table 1.1. Figures 1.10, 1.11, and 1.12 show the HiSPARC

observations around the epochs of the first four FDs listed in Table 1.1. 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. 1.4).

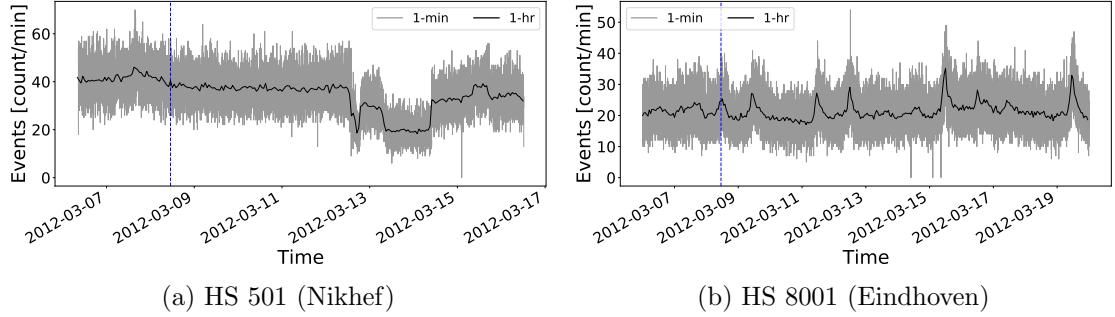


Figure 1.10: HiSPARC data for stations 501 and 8001 around the epoch of the FDs in March 2012. The plot shows the minute-averaged and hourly-averaged trigger events between detectors within the station. The vertical blue-dashed lines show the approximate onset-time of the FDs. The units of time on the x-axis are, YYYY-MM-DD.

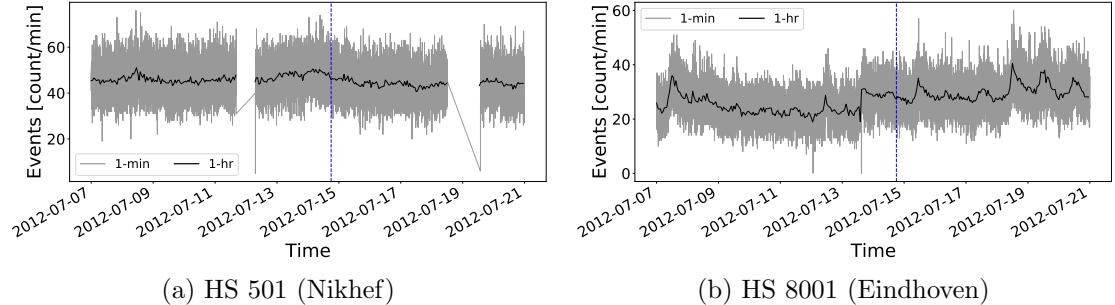


Figure 1.11: HiSPARC data for stations 501 and 8001 around the epoch of the FD in July 2012. The plot shows the minute-averaged and hourly-averaged trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

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 1.10a); however, it is unclear whether this was a consequence of the FD or other hardware 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

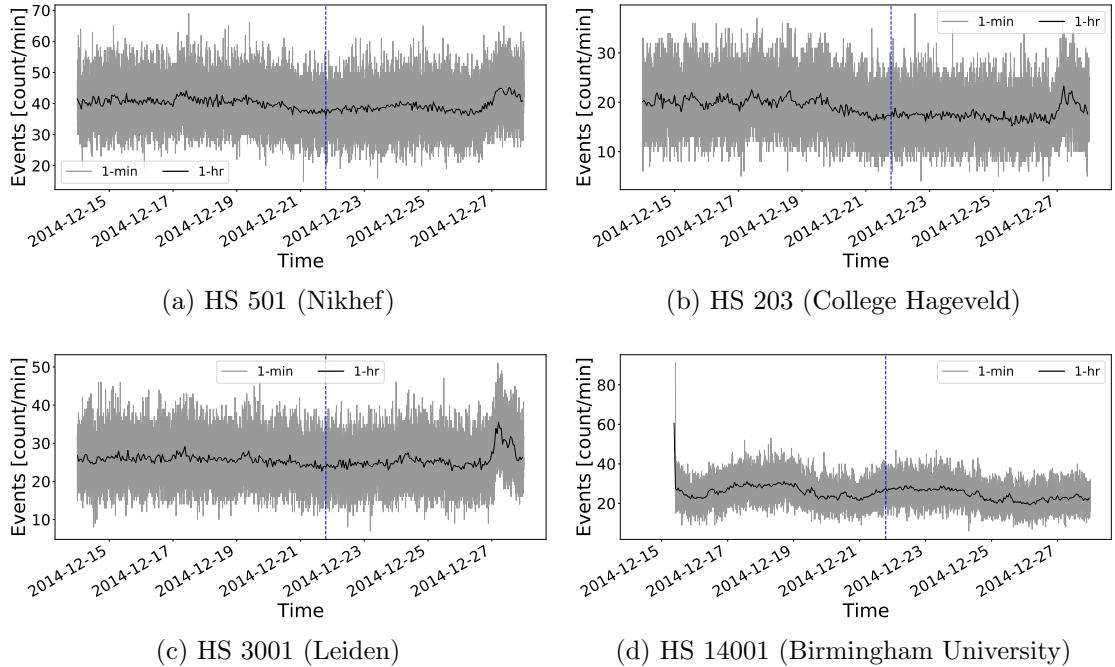


Figure 1.12: HiSPARC data for four stations around the epoch of the FD in December 2014. The plot shows the minute-averaged and hourly-averaged trigger events between detectors within the station. The vertical blue-dashed line shows the approximate onset-time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

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 variations in the atmospheric pressure. In Section 1.6 this is discussed further and the effect is accounted for.

In addition, it is quite clear from Figure 1.10b and Figure 1.11b that station 8001 (Eindhoven) displays a semi-persistent diurnal variation in the count rate. We know that daily variation is a combination of the CR anisotropy in the interplanetary space and a consequence of the rotation of the Earth meaning detectors look in different directions over the course of a day; however, we typically expect this diurnal variation to be $< 1\%$ ([Mishra & Mishra, 2007, 2008](#); [Dubey et al., 2016](#); [Thomas et al., 2017](#)), but here we see an increase of $\sim 100\%$, which suggests there may be an additional factor due to another source, likely originating in hardware.

For the final two FDs listed in Table 1.1, the plot of the HiSPARC observations is shown in Figure 1.13. 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 $\sim 50\%$ (see detector channel 2 in Fig. 1.13c). 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 1.13c 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 in the next section we remove these effects to standardise the HiSPARC data.

1.6 Atmospheric Corrections of HiSPARC Data

1.6.1 Motivation

It is well known that observations made by ground-based CR detectors are susceptible to atmospheric conditions ([Dorman, 2004, 2010](#); [Berkova et al., 2011](#); [De Menonça et al., 2013](#); [Paschalis et al., 2013](#)). As we have seen in the plots shown in Section 1.5, there exist excursions in the data whose origins are from atmospheric

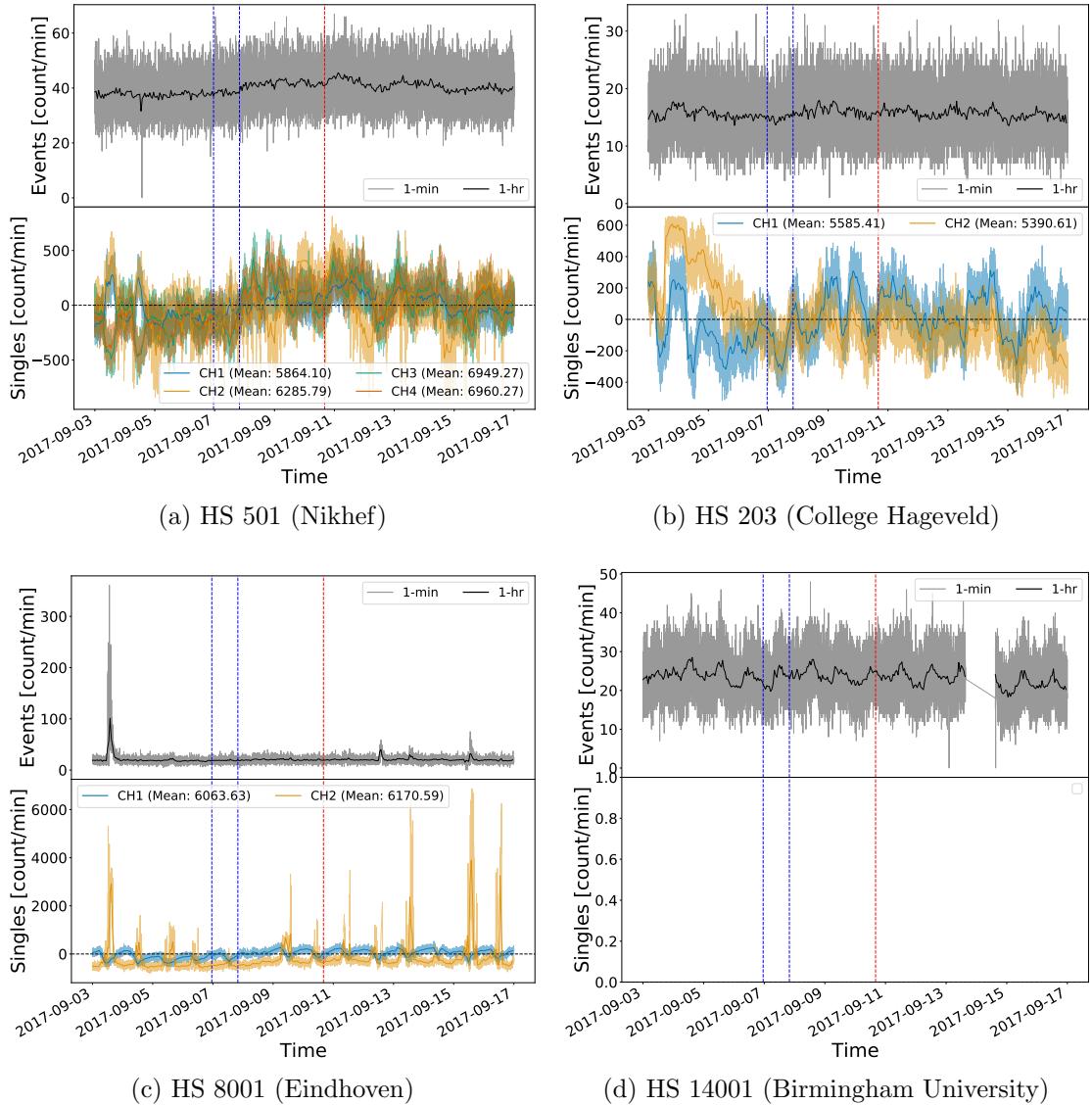


Figure 1.13: HiSPARC data for four stations around the epoch in which there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute- and hourly-averaged trigger events between detectors within the station, while the bottom panel shows the minute- and hourly-averaged 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.

variations. This sensitivity makes it difficult to differentiate between variations due space weather events and those due to Earth's atmospheric conditions, therefore it was necessary to correct for these effects in the data.

1.6.2 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 NM and MDs in Figure 1.14.

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.

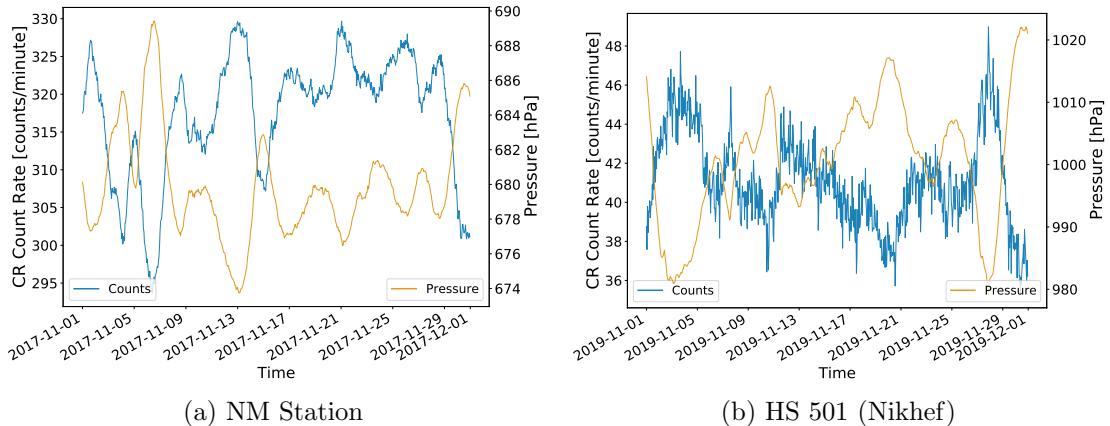


Figure 1.14: The anti-correlation between CR count rates and the atmospheric pressure. (a) shows the CR and the local atmospheric pressure measured at a NM in the South Pole; (b) shows the CR and pressure measured by HiSPARC station 501. In both plots, the data shown are hourly-averaged, to highlight the effects.

The 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 CR variations (N) that depend on the local atmospheric pressure are described by:

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

$$\Delta N = -\beta N \Delta P, \quad (1.2)$$

where ΔN is the change in 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 (1.2), the solution shows the dependence of cosmic ray intensity on pressure,

$$N = N_0 e^{-\beta \Delta P}. \quad (1.3)$$

Therefore by taking the logarithm of equation (1.3), one can obtain the barometric coefficient by fitting the linear model to the observed data, of the form:

$$\ln \left(\frac{N}{N_0} \right) = -\beta \Delta P, \quad (1.4)$$

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 1.15. 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.

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 1.16. It is clear from Figure 1.16 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 1.17 for monthly corrections throughout 2017 for the NM station at the South Pole (SOPO).

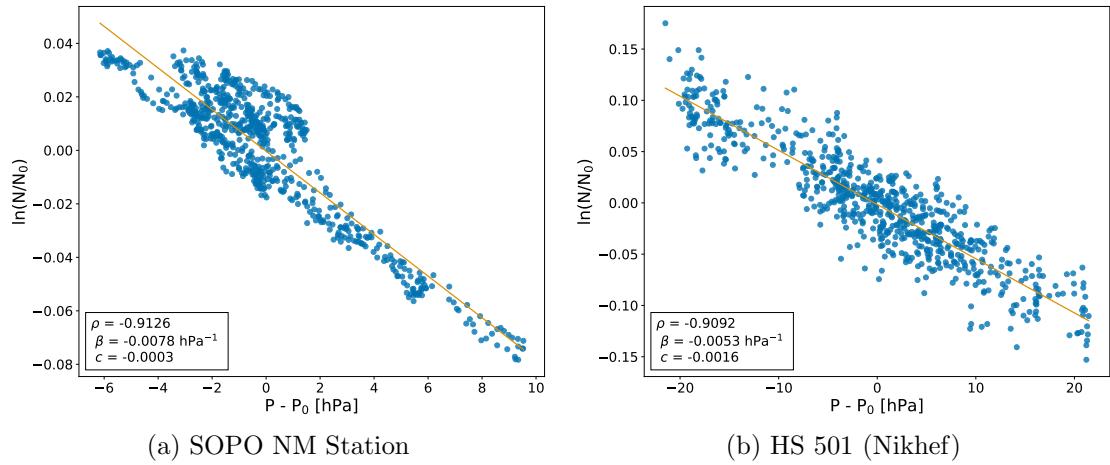


Figure 1.15: The barometric coefficient calculation: (a) during November 2017 for the South Pole (SOPO) NM station, (b) during November 2019 for HiSPARC station 501 at Nikhef.

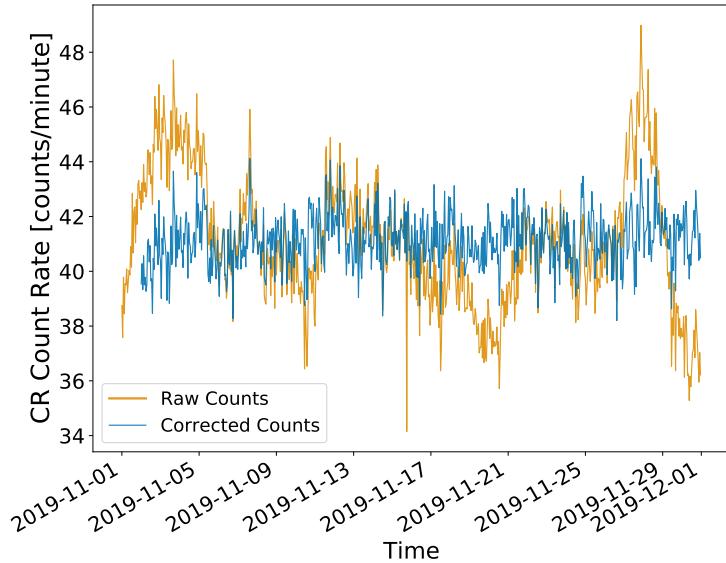


Figure 1.16: A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the pressure correction (blue line).

Figure 1.17 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 (Apatity and Rome), thus validating that the method used in this work was suitable for application on the HiSPARC data. The barometric correction was performed on the HiSPARC data for stations where sufficient pressure data and count rates existed.

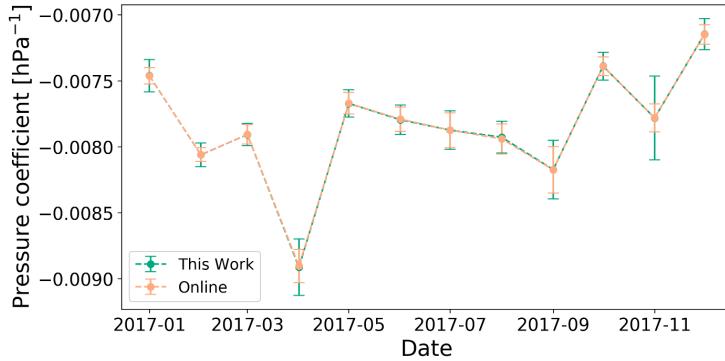


Figure 1.17: A comparison between the monthly barometric coefficient computed in this work and using the online barometric coefficient tool throughout the year 2017 for the SOPO NM station.

1.6.3 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. 1.9c). However, when observing the singles rates, we do see some detectors displaying a positive effect and some displaying a negative effect (see Fig. 1.9). 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 long 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; hence 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 located. The latter is not of use, but the former can be used for temperature corrections. Figure 1.18a 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 1.18b, yet weaker than with pressure.

We simplified the correction method and used a linear fitting technique (i.e. the same as the barometric correction in Section 1.6.2; however, replacing the pressure for locally measured outdoor temperature). We can see from Figure 1.19 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.

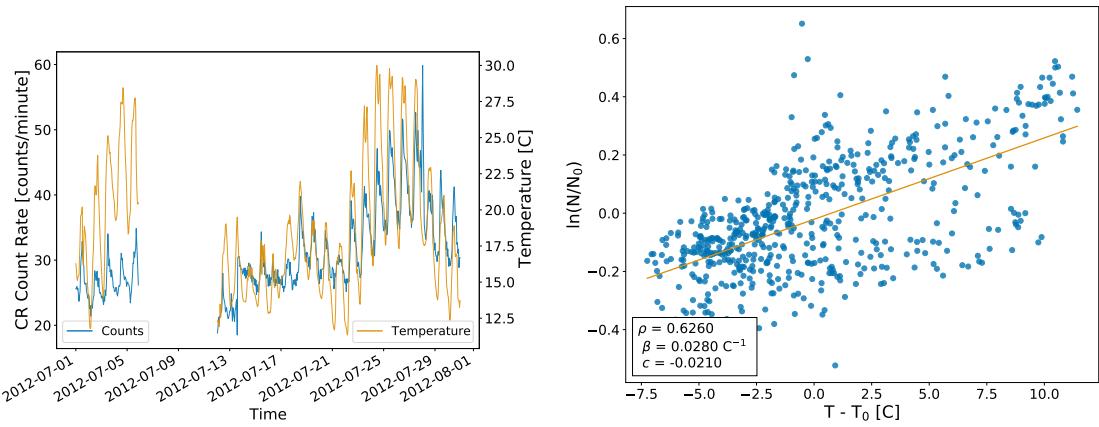


Figure 1.18: The relationship between the pressure corrected events data and the outdoor temperature as measured at HiSPARC station 8001 (Eindhoven). (a) shows the time-series of hourly-averaged pressure corrected events and temperature data; (b) shows the correlation between the counts and temperature, and the fitted line to calculate the correction coefficient.

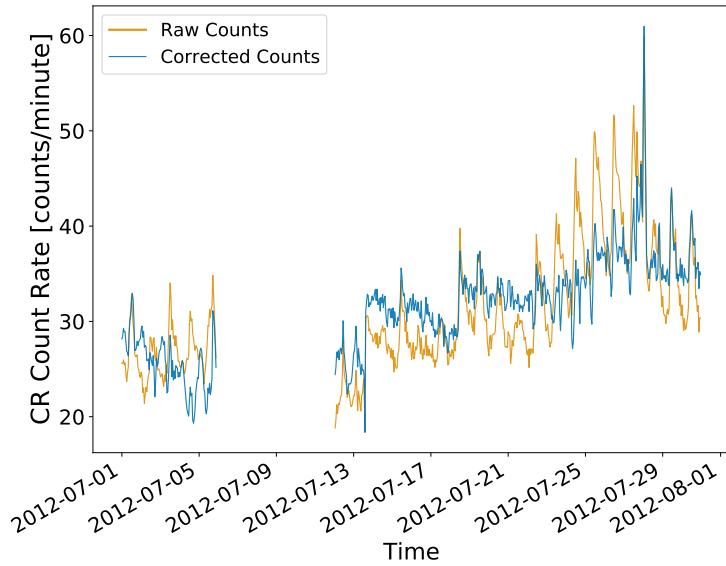


Figure 1.19: A comparison between the hourly-averaged HiSPARC count rate before (orange line) and after the temperature correction (blue line). After the correction, the diurnal variation from the temperature effect has been reduced.

1.7 HiSPARC Observations After Atmospheric Corrections

1.7.1 Observations of Ground Level Enhancements

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 1.20 and Figure 1.21 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.

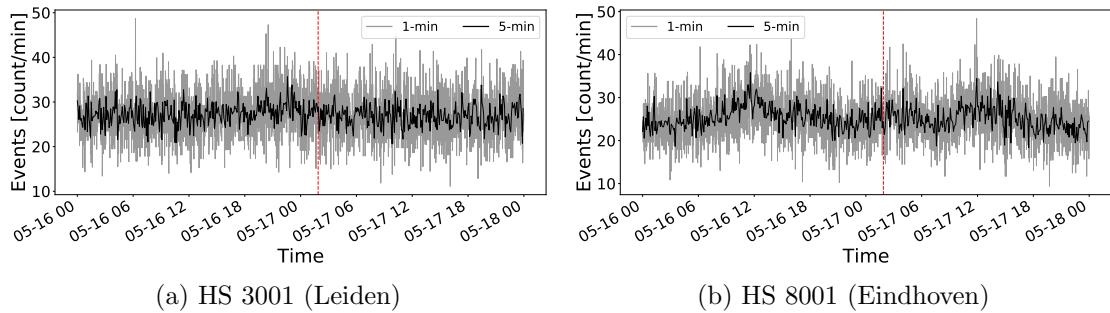


Figure 1.20: Atmospheric-corrected HiSPARC data for stations 8001 and 3001 around the epoch of GLE 71 on 17/05/2012. The plot shows the minute-averaged and 5-minute-averaged trigger events between detectors within the station. The vertical red, dashed line depicts the approximate onset time of the GLE. The units of time on the x-axis are, MM-DD HH.

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 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. \(2012\)](#) stated that NM

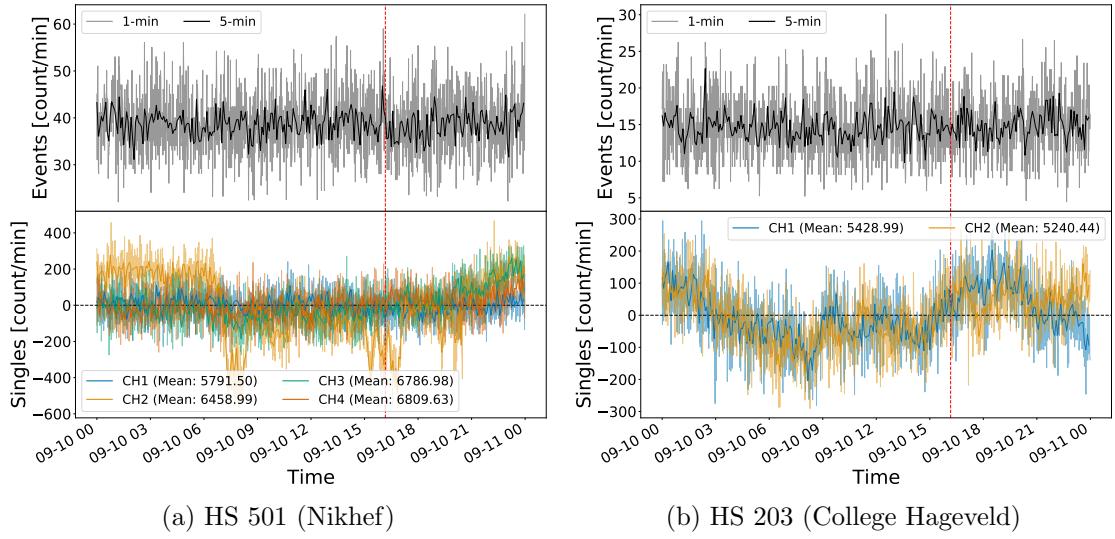


Figure 1.21: Atmospheric-corrected HiSPARC data for 2 stations around the epoch of GLE 72 on 10/09/2017. The top panel of each subplot shows the 1- and 5-minute averaged trigger events between detectors within the station, while the bottom panel shows the 1- and 5-minute averaged singles counts, mean-subtracted, 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.

stations with cut-off rigidities up to ~ 15 GV observed the GLE in September 1989, indicating that high-energy SEPs may be present and 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 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 instead of 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 1.8.

1.7.2 Observations of Forbush Decreases

The search for evidence of FDs was also re-conducted using the atmospheric-effect corrected HiSPARC data. Figure 1.22 and Figure 1.23 shows the corrected HiS-

PARC observations around the epochs of a FD in July 2012 and December 2014, respectively.

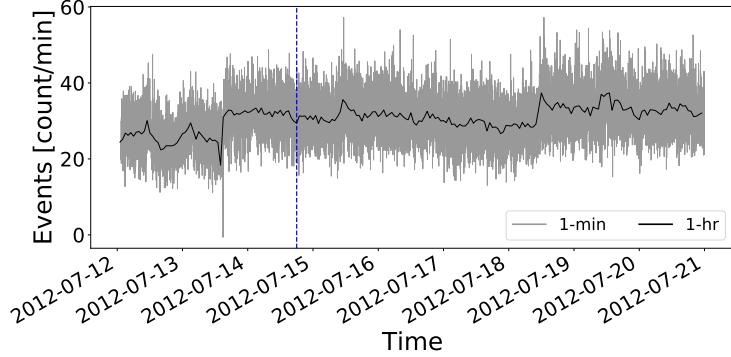


Figure 1.22: Atmospheric-corrected HiSPARC data for station 8001 (Eindhoven) around the epoch of the FD in July 2012. The plot shows the minute- and hourly-averaged trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

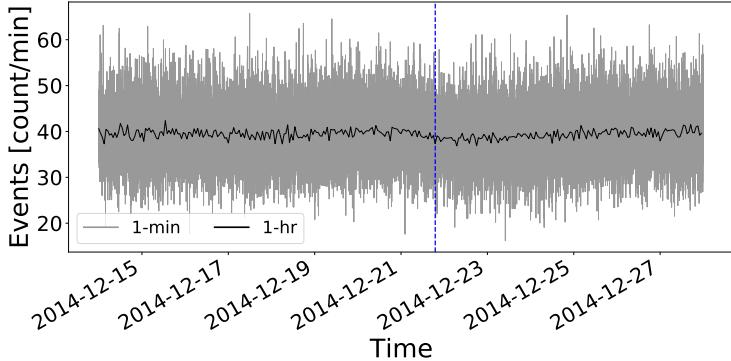


Figure 1.23: Atmospheric-corrected HiSPARC data for station 501 (Nikhef) around the epoch of the FD in December 2014. The plot shows the minute- and hourly-averaged trigger events between detectors within the station. The vertical blue, dashed line depicts the approximate onset time of the FD. The units of time on the x-axis are, YYYY-MM-DD.

In Figure 1.22 there are no clear FD observations in the corrected HiSPARC data; however, in Figure 1.23 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 the HiSPARC network.

Finally, the atmospheric-effects corrected observations of the two FDs which occurred around GLE 72 is shown in Figure 1.24, showing the corrected events data

and singles data for each of the individual detectors in the stations.

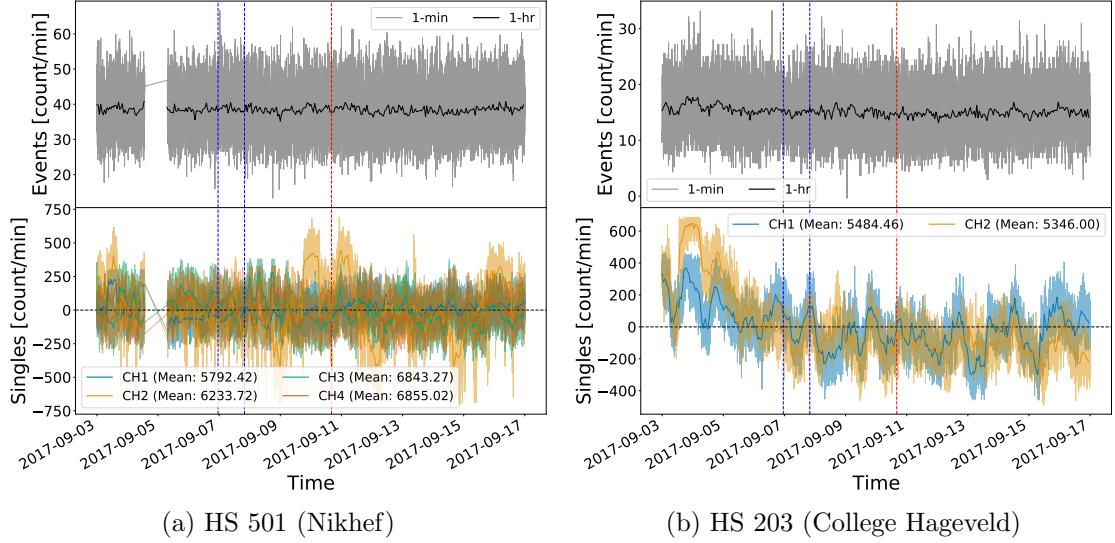


Figure 1.24: Atmospheric-corrected HiSPARC data for 2 stations in an epoch where there were two FDs close to the onset of GLE 72. The top panel of each subplot shows the minute- and hourly-averaged trigger events. The bottom panel shows the minute- and hourly-averaged singles counts, mean-subtracted, 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 1.24, 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.

1.8 Air Shower Simulations

1.8.1 Motivation

There is 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 PCRs. To investigate this, simulations of air shower development were performed for a range of PCRs 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 PCRs 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 – 10^{20} eV, and 4×10^9 – 10^{20} eV, respectively. In total $\sim 2 \times 10^5$ proton-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.

1.8.2 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 1.25 shows the radial distribution of muons at ground level for air showers induced by vertically incident protons and α -particles.

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 1.25, but they are not shown here as the difference is not drastically different by-eye.

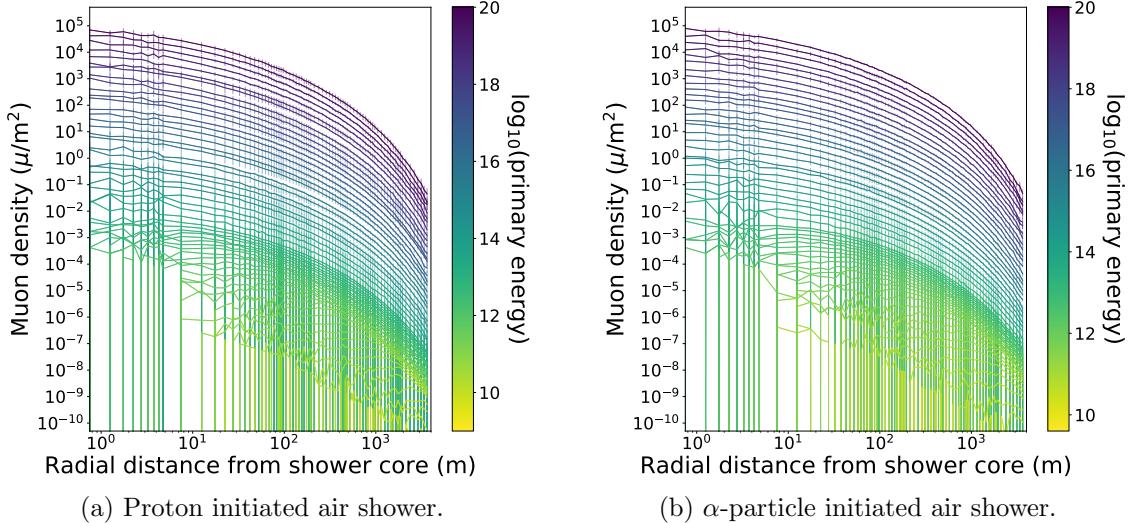


Figure 1.25: Mean muon density footprints for (a) proton-initiated air showers and (b) α -particle-initiated air showers with initial PCR trajectories with zenith angles $\theta = 0^\circ$ and various PCR energies. The error bars given represent 1σ .

The interpretation of Figure 1.25 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. The simulations show, to trigger two HiSPARC detectors separated by ~ 10 m by muons from the same shower, the density of the shower needs to be at least $\sim 0.005 \mu\text{m}^{-2}$, suggesting 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. \(2020\)](#).

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., 2020](#)). 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.

1.8.3 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 1.26 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 1.26, but they are not shown here, as the difference is again not drastically different by-eye.

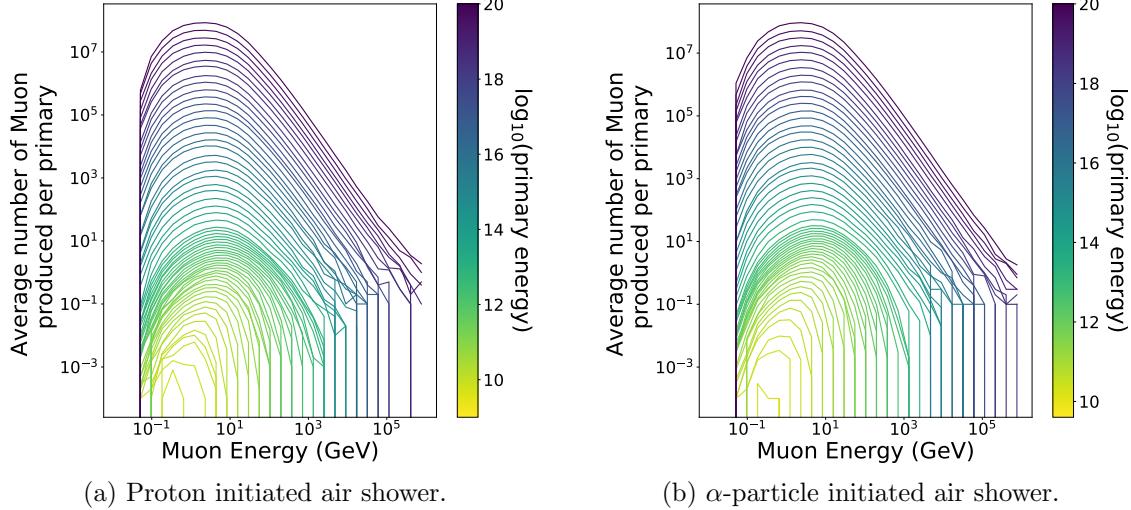


Figure 1.26: Mean number of muons produced at ground level by the PCR for (a) proton-initiated air showers and (b) α -particle-initiated air showers, for various PCR energies.

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. energies $< 10^9$ eV, this helps explain why the space weather events were not observed in the HiSPARC events data. The 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. Fig-

Figure 1.27 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 1.27, we see that the ground-based flux is similar for both types of simulation performed. In both, the low-energy muon flux dominates, and peak at a muon energy of ~ 1 GeV.

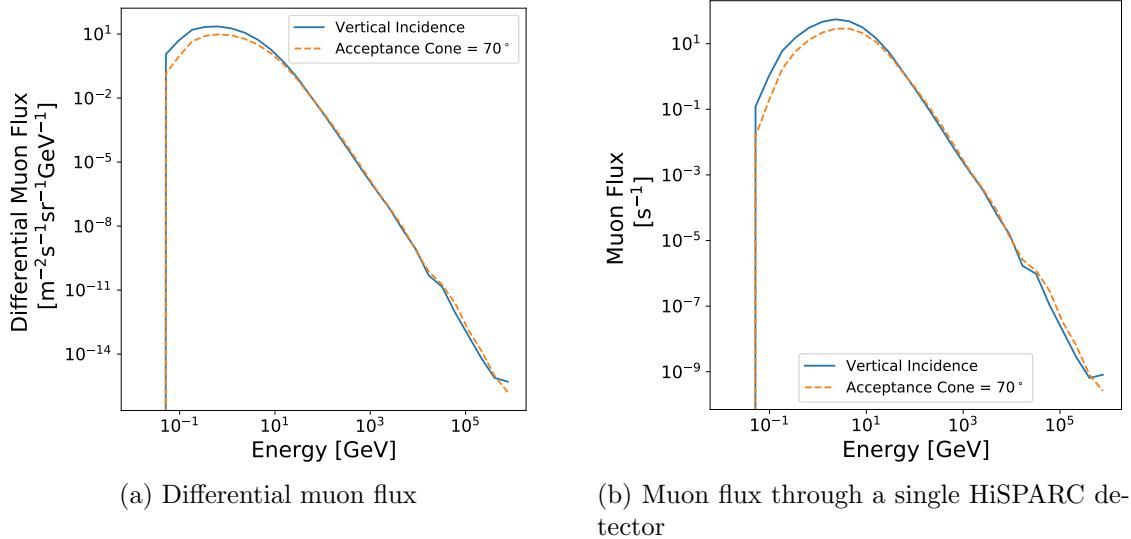


Figure 1.27: Ground level muon spectra as computed in the CORSIKA simulations. (a) shows the differential muon flux at ground level; (b) shows the muon flux through a single HiSPARC detector. In both plots the solid, blue line shows the simulations using vertically incident PCRs, and the dashed, orange 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: $\sim 85 \mu/\text{s}$ (for non-vertical, i.e. 70° acceptance cone simulations), and $160 \mu/\text{s}$ (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., 2020](#); [Particle Data Group et al., 2020](#)), or $1 \mu \text{ cm}^{-2} \text{ min}^{-1}$ ([Particle Data Group et al., 2020](#)).

To understand, from the CORSIKA simulations, how the count rate changed 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.

1.8.4 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. 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. The GLEs within the MAIRE tool are detailed in Table 1.3, also providing an estimate for the maximum increase in count rate, as observed in the NMDB, as a reference.

Table 1.3: 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	29/29/1989	~ 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)

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 1.1. Nevertheless, the MAIRE simulations helped in our understanding of whether it was likely or possible to observe a GLE using data 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 1.28 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).

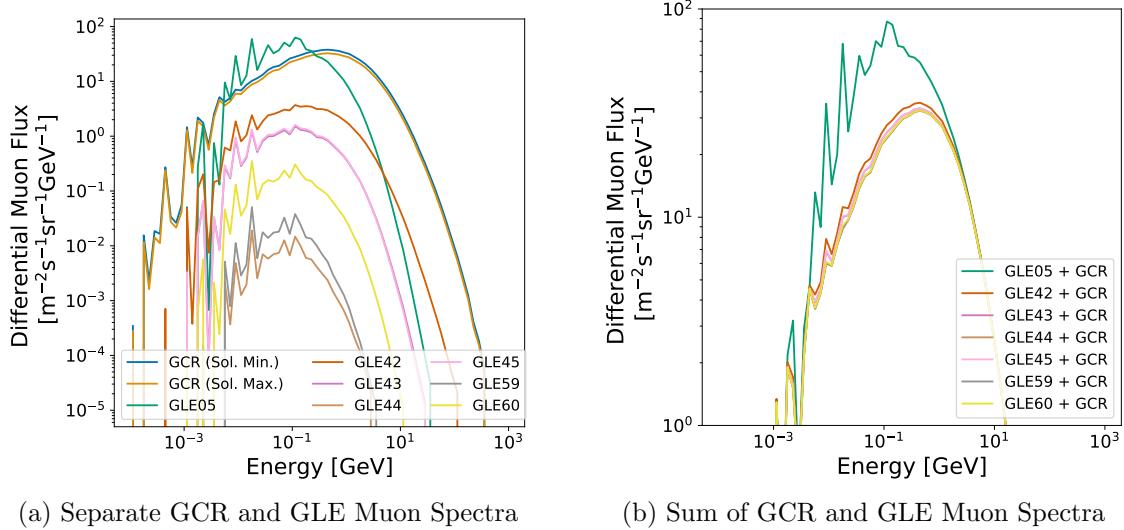


Figure 1.28: Plots of the calculated MAIRE muon spectra for different incident PCR spectra. Blue and orange lines show muon spectra calculated for the incident GCR spectra during solar minimum and maximum, respectively. The other coloured lines show the computed muon spectra for the incident GLE spectra. (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.

The GCR-induced muon spectrum in Figure 1.28 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 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 NM 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](#)).

Table 1.4 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). The muon count rate increases predicted by MAIRE are also shown in Table 1.4, for comparison.

Table 1.4: Observed and predicted increases in the CR count rates at the Kiel NM station, with rows ordered by the observed NM 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	-	-	-	-	-

We can see from Table 1.4 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. More likely, the difference is due to the anisotropy of the CRs initiating the GLEs, as MAIRE does not adequately account for anisotropy in the simulations ([Dyer et al., 2003; Lei et al., 2004](#)). The ranked ordering from strongest to weakest GLEs in the observed Kiel data and that predicted by MAIRE are in good agreement, suggesting that despite an incorrect

amplitude of the increase, generally the two data sets agree on which GLEs were the strongest. The ratio column gives a factor to convert from the MAIRE predicted increase to the NMDB observed increase. This was used as a calibration factor, to ensure agreement between the MAIRE predictions and the observed data. Furthermore, as the ratio provides us with a calibration to recover the observed increase in neutron counts from the MAIRE simulations, we assumed the same calibration factor may be applied to the predicted increase in the muon counts. This conversion is shown in Table 1.5.

In addition, comparing the fraction of the predicted muon to neutron increase shown in Table 1.4, we see the significantly reduced prediction for the muon count increase. We know this is due to the energy spectrum of the PCRs, but the values in Table 1.4 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.

We calculated the increase in muon count rate at several HiSPARC stations, and the predicted rates were adjusted by the ratio factor in Table 1.4. These calculations relied on the assumption that the 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. The predicted count rates, adjusted for the factors in Table 1.4, are shown in Table 1.5. Again, GLE 5 has been omitted from these results as the Kiel station was not online then, so we cannot calibrate the MAIRE predictions.

The predicted increase in the muon count rates, calculated using the MAIRE simulations, for the NM station and the HiSPARC stations, in both the Netherlands and the UK, are in near-agreement. Assuming the effects of anisotropy are small, this suggests that the predicted spectra at the HiSPARC stations (adjusted for the calibration factor) are representative of the approximate increase one would expect to observe.

The effect of the GLEs shows only a small increase in the muon count rate,

Table 1.5: 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 1.4. The rows ordered by the observed NM magnitude from Table 1.4.

GLE	Predicted 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%
5	-	-	-

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 1.1, suggest that we would have expected an increase in the muon spectrum of $< 0.1\%$ for both GLE 71 and 72, which rules 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 1.4 and Table 1.5 suggests we would still only expect an increase in the HiSPARC count rate between $\sim 0.1 - 2\%$. 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.

1.9 Conclusion

We have presented a feasibility study on using the existing HiSPARC network of muon detectors to monitor space weather events. This was performed 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 removed from the data.

After the removal 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

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 Air Showers (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. 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, $R_C \sim 3.0 - 3.5$ 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 only suitable as a monitor of the rarer and 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 major issues with the HiSPARC network in its current configurations is the design of the stations. The triggering of multiple detectors biases the stations to observe higher energy PCRs. Since 2016, several HiSPARC stations have been providing the count rates of the individual detectors (the singles rates), which overcomes this biasing problem; however, these data are often inconsistent between stations and

also show a strong diurnal variation that we believe is the manifestation of thermal noise.

The current method of removing the temperature-induced diurnal variation uses the atmospheric temperature, and not the temperature inside of the rooftop boxes, as it is the only temperature data available. We believe 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.

2 HiSPARC Station 14008

2.1 Introduction

It was shown in Chapter 1, using data acquired from the High School Project on Astrophysics and Research with Cosmics (HiSPARC) network, that it was not clearly capable of observing space weather events. In part, we showed that this was due to the low-magnitude of the increase in muon flux during such events, but it was because of the current configuration’s bias towards higher energy Cosmic Rays (CRs), and finally because the detectors are rather sensitive to variations in the terrestrial 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 be not always so efficient, as different detectors in the HiSPARC network displayed different responses to pressure and temperature variation. In addition we showed that the relationship between atmospheric temperature and CR count was non-trivial, unlike the counterpart correction for pressure.

Thermal fluctuations in the atmosphere clearly induce thermal noise in the Photo Multiplier Tubes (PMTs) and although the temperature inside the HiSPARC roof boxes have 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 1 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 explored the option of designing an alternate configuration of HiSPARC station such 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, it was possible to stack two detectors on top of each other to measure a single muon which traverses both scintillators, hence inducing signals in both PMTs.

2.2 Aims

The aim of creating a new HiSPARC station was to test whether an alternative configuration of a HiSPARC station could minimise atmospheric variations in the data, and to standardise 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 any space weather events with the alternative HiSPARC configuration.

2.3 HiSPARC Station 14008 Set-up

2.3.1 Configuration

The configuration of HiSPARC station 14008 is shown in Figure 2.1; the station is composed of two detectors stacked on top of each other, both inside one roof box. This configuration is advantageous, over the single scintillator, single PMT, HiSPARC set-up, as it allows the recording of single muons which traverses both scintillators.

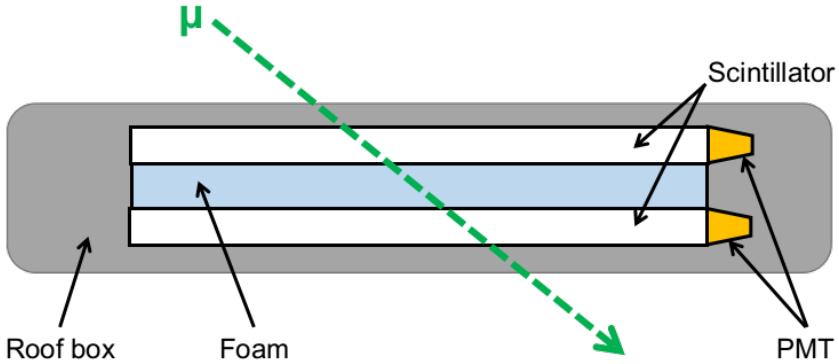


Figure 2.1: Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box.

We showed in Chapter 1 that the existing HiSPARC design requiring coincident 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 in Figure 2.2a. Upon the completed assembly of the detectors, they were placed within the roof box on the roof of the Poynting Physics building on the campus of University of Birmingham.

Propagating charged particles lose energy in matter. Derived from the Bethe-Bloch formula, we can estimate the amount of energy lost by a particle in a material given by:

$$\Delta E = \Delta x S \rho \cos(\theta), \quad (2.1)$$

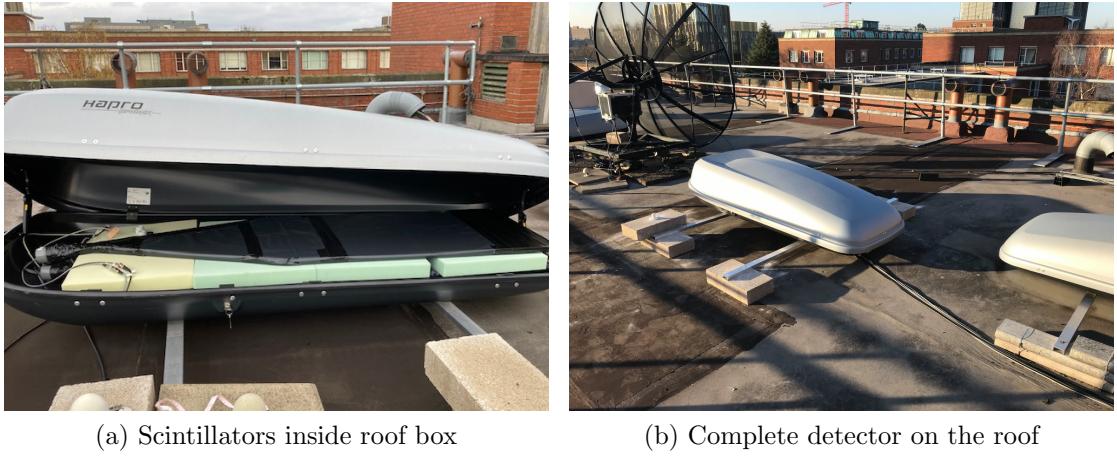


Figure 2.2: HiSPARC 14008 assembly and configuration. (a) Shows the stacked arrangement of the scintillators within the roof box, between layers of protective foam. (b) Shows the complete detector on the roof of the Poynting building on the University of Birmingham campus.

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. \(2020\)](#) states the most probable energy loss of a vertically incident muon in a single scintillator is 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 ~ 7.42 MeV. This new lower limit does not significantly change the values of the predicted GLE magnitudes in Section 1.8.4.

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, ~ 70 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 full 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 2.3.

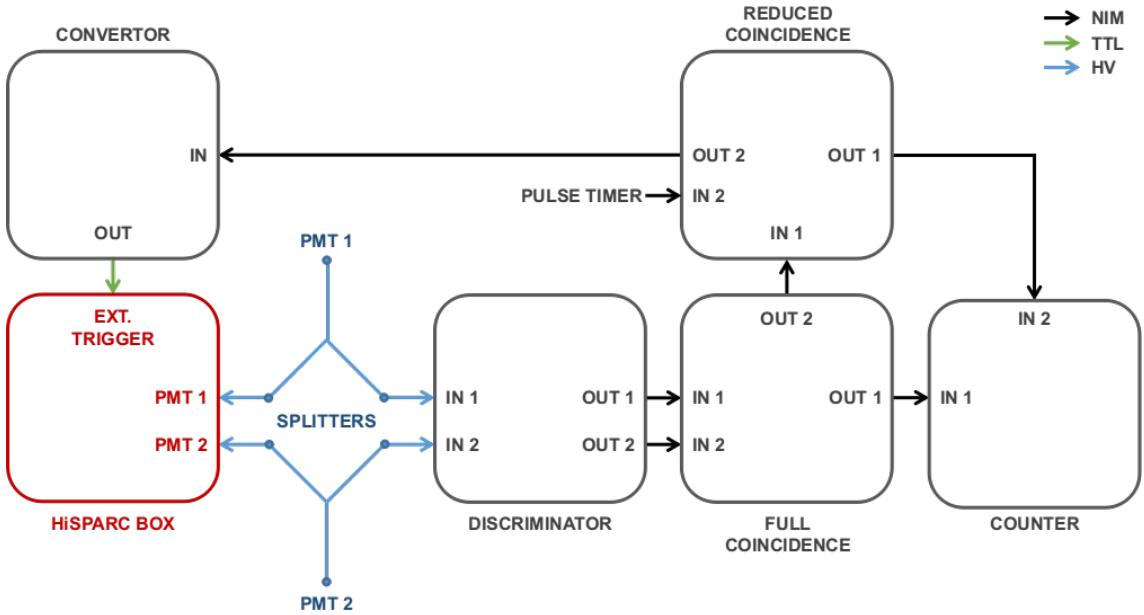


Figure 2.3: Schematic diagram of the HiSPARC 14008 station NIM crate configuration. Black-outlined boxes indicate the NIM crate modules, while red-outlined boxes depict HiSPARC hardware modules. Black arrows depict a NIM signal; green arrows show a TTL signal; blue arrows depict a HV signal. The PMTs from the detectors 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 2.3.4, but here we discuss the configuration of the NIM crate set-up. The High Voltage (HV) 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 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 HV sig-

nal 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 full 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 full coincidence signal and the pulse timer gate ensures that the full 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) convertor and the output from this is used as an external trigger signal to trigger 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, but with a reduced frequency by a factor of ~ 100 .

The use of these NIM modules introduce 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](#)). The introduced delays between the PMT output, to being registered at different end-points, due to NIM modules and cables are listed in Table 2.1.

From Table 2.1 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 a $\sim 8\ \text{ns}$ 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 2.1: 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 this is due to the low muon count rate and the wide pre- and post-trigger windows of the HiSPARC data acquisition software. [Fokkema \(2012\)](#) shows the HiSPARC data acquisition software uses pre-trigger (1 μ s), coincidence (1.5 μ s) and post-trigger (3.5 μ s) windows. This means that the station is sufficient to measure signals with delays of tens of nanoseconds, as they will be captured within the 1 μ s pre-trigger window and thus counted. Using the NIM counter, as is discussed in Section 2.3.4, we measure all counts on a cadence of 10–seconds, therefore a delay of \sim 8 ns does not impede our ability to count the same events during the cadence.

2.3.2 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 voltage. Each of the detector PMTs needs to be powered with a high enough operating voltage such 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 2.4, 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 2.4. The voltage calibration plot shows drastically the different performances one can get from different PMTs, therefore it was necessary to treat each PMT individually when calibrating.

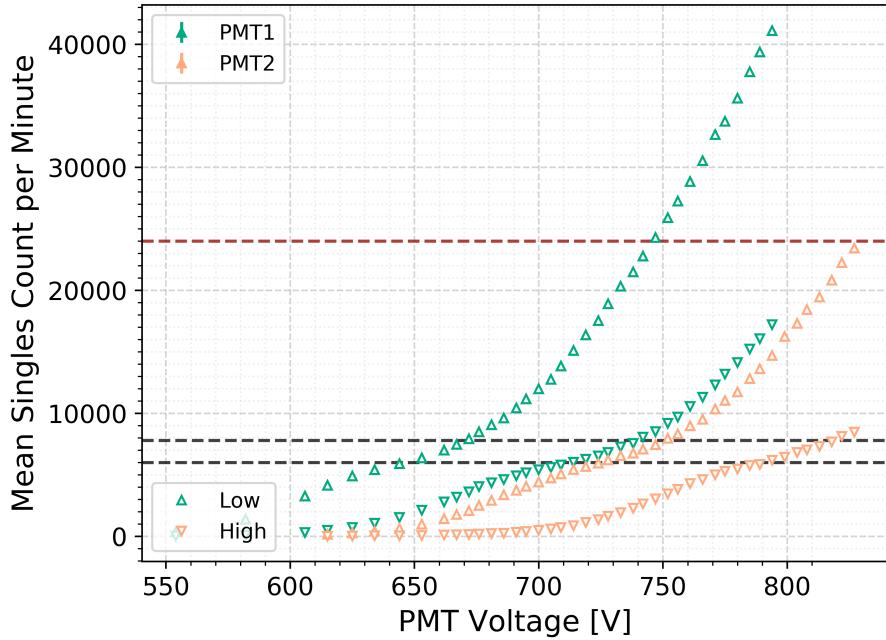


Figure 2.4: Voltage calibration curve for the PMTs of station 14008. The upper, red-dashed line indicates the upper limit for the low threshold singles rate (400 Hz), and the lower 2, black-dashed lines indicate the upper and lower bounds for the high threshold singles rate (100–130 Hz).

Initially the station was set-up supplying PMT1 and PMT2 ~ 725 V and ~ 790 V, respectively, based on the calibration in Figure 2.4. 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.

2.3.3 Monitoring Temperature

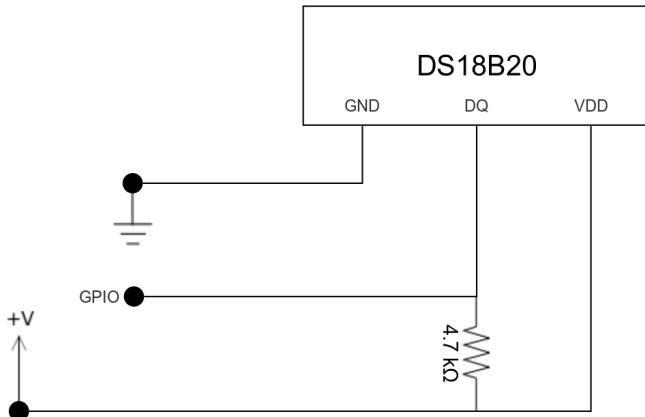
In Chapter 1, we suspected that the singles count rates were affected by the temperature of the PMT within the HiSPARC roof-boxes. Some of the existing HiSPARC stations monitored local atmospheric temperature, however none measured the temperature of the PMT inside the roof box. When building this new HiSPARC station, a temperature sensor was placed into the roof box which allowed us to monitor the temperature of the PMT more accurately.

Figure 2.5a shows the circuit diagram for the temperature sensor. 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; the microcontroller used was a Raspberry Pi 4 (see Section 2.3.4). 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.

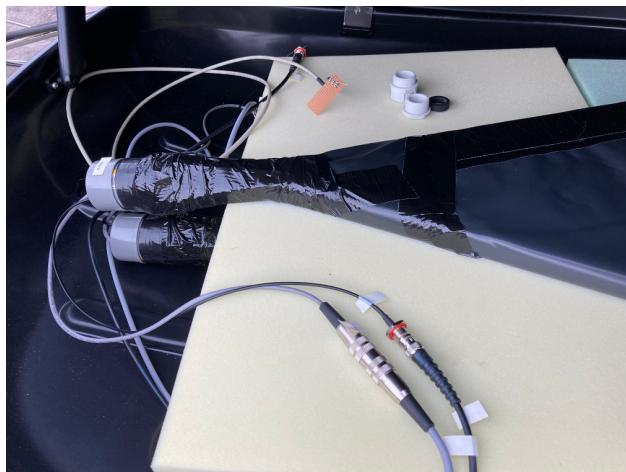
2.3.4 Data Acquisition

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 full 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 full 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 2.6.

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 data from the NIM counter and the temperature data from the DS18B20 sensor to local files on the Raspberry Pi. Both the coincidences and temperature data are recorded on a 10



(a) Circuit diagram



(b) Sensor within roof box

Figure 2.5: (a) Schematic diagram of the DS18B20 temperature sensor circuit, whereby the voltage, ground, and GPIO interfaces connect directly into 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.

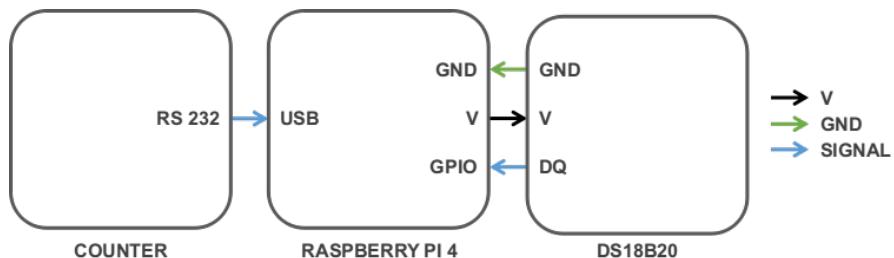


Figure 2.6: Schematic diagram of the HiSPARC station 14008 data acquisition interfaces. Black arrows depict a 5 V signal; green arrows show ground interfaces; blue arrows depict the 1-wire signal output from the temperature sensor.

second cadence.

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 2.2.

Table 2.2: 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 2.3.

Table 2.3: 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 1.3.

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. In these early stages the data are unreliable, so it is not recommended to use this data. 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 they are available for both coincidences and temperature, are regularly available, and were acquired in the configuration described above.

2.3.5 Monitoring Pressure

As with 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

¹Disk location: /disk/moose/general/epesv001/datadisk/147.188.46.117.hisparc_pi/

in Coleshill, Warwickshire (ID: 19187), nearby Birmingham International Airport, ~ 20 km from the HiSPARC detectors; we believe the pressure variation is small over this distance. The pressure is measured at the MIDAS station level and a correction for altitude is not applied.

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.

2.4 Methodology

2.4.1 Atmospheric Corrections

After considerate review of the methods for temperature correction, the method discussed and used in Chapter 1 was used here, i.e. using a linear relationship between CRs and temperature; 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 CRs and temperature (defined by equation (1.4), changing pressure to temperature) the long-term temperature relationship was fit and was removed from the data.
- After removing the long-term temperature relationship, for each day the temperature correction was applied again to remove the daily variations. During the fit to the data the values of N_0 and T_0 used were the de-trended, 24-hr smoothed data.

This was done, as 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, without

using the de-trended, 24-hr smoothed data as the value for the values of N_0 and T_0 , we found that after the corrections, there existed jumps between days; this method ensured that there was a smooth transition in the corrected data between days.

Using the theory outlined in Section 1.6.2 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 (1.4). Similarly, the barometric correction was performed over durations of 1-month at a time.

For reasons discussed in Section 2.5 the temperature correction was, in practise, only applied to the singles data. When correcting the atmospheric effects on the singles data, the temperature correction process was performed first and was followed by the pressure correction process.

2.4.2 Observations

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

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (2.2)$$

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 with 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:

$$P(n; \lambda_1, \lambda_2) = P(n; \lambda_1 + \lambda_2), \quad (2.3)$$

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

Using this information, it was possible to use a sampling algorithm to determine the mean level and noise of the HiSPARC 14008 station's data. 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.

To test the station's performance 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 can use the fact that count data follow a Poisson distribution to compute the probability of statistically significant spikes in the data. The Poisson cumulative distribution function is given by:

$$F(k; \lambda) = \sum_{i=0}^k \frac{\lambda^i e^{-\lambda}}{i!}, \quad (2.4)$$

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

$$p(k) = 1 - F(k - 1; \lambda). \quad (2.5)$$

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 cadence 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 is:

$$p_N = 1 - [1 - p(k)]^N, \quad (2.6)$$

where a low value for p_N indicates that the observed event is considered statistically significant.

This can be generalised using the cumulative binomial distribution ([Basu & Chaplin, 2017](#)). The probability of finding at least r occurrences in N -cadences at or above k , given the mean level, λ , is given by equation (2.7), which is equal to equation (2.6) when $r = 1$,

$$p[r; p(k), N] = \sum_{r=r}^N \binom{N}{r} p(k)^r [1 - p(k)]^{N-r}. \quad (2.7)$$

By applying equation (2.7) to the data, we were able to test whether there were any significant events. Again, a low value for $p[r; p(\lambda), N]$ indicates that the event is statistically significant.

Another frequentist test that was used makes use of the assumption 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:

$$s = k - \lambda, \quad (2.8)$$

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:

$$Z = \frac{s}{\sigma}, \quad (2.9)$$

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.

As a further test we also analysed re-sampled versions of the artificial data. This allowed us to reduce the Poisson noise, with the ambition of reducing the noise such to allow us to observe lower amplitude space weather events. We also ran the statistics tests on 1-minute and 5-minutes averages of the artificial data. The

statistics tests were run under the same underlying principals, however instead of using the Poisson cumulative distribution in equation (2.5), 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. Similarly, using the Gaussian distribution with mean, $\mu = \lambda$, and standard deviation, $\sigma = \sqrt{\lambda/n}$, in equation (2.9).

2.5 Atmospheric Corrections

2.5.1 Temperature Correction

Using the method outlined in Section 2.4 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 2.7.

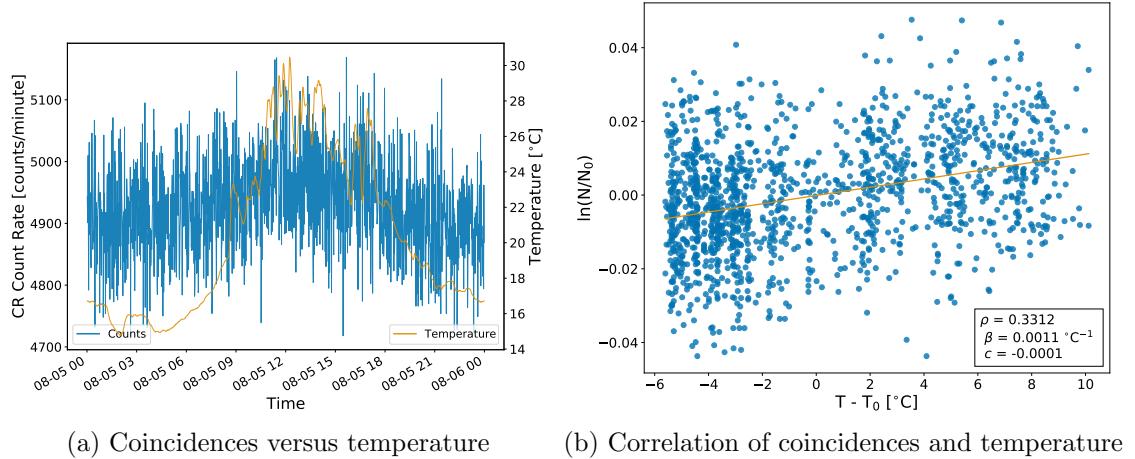


Figure 2.7: The relationship between the full coincidences data and the temperature within the roof box over a single day. (a) Shows the comparison between coincidences data (blue) and temperature data (orange); (b) shows the correlation between the coincidences counts and temperature, and the fitted line to calculate the correction coefficient.

We see from Figure 2.7 that there is a weak correlation between the temperature within the roof box and the CR counts. Noting the common adage, that correlation does not necessarily imply causation, is important here in this case of weak correla-

tion. We believe that the relationship between the coincidences and the temperature in the roof box is likely not 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 2.6.1 that the PMT thermal noise, which has a strong diurnal component, does not bleed through into the noise on the coincidences data. We believe 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. 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 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 2.8.

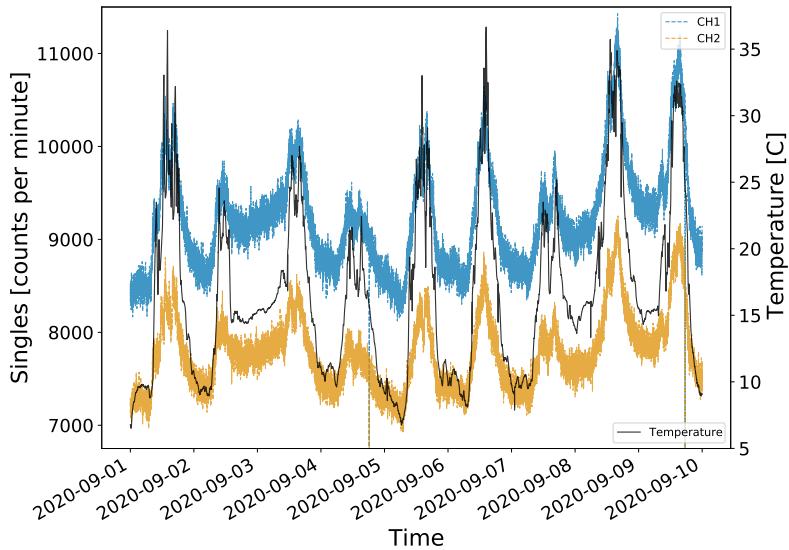


Figure 2.8: The relationship between the full singles data (blue and orange lines) and the temperature within the roof box (black line).

Figure 2.8 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 1 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 2.9.

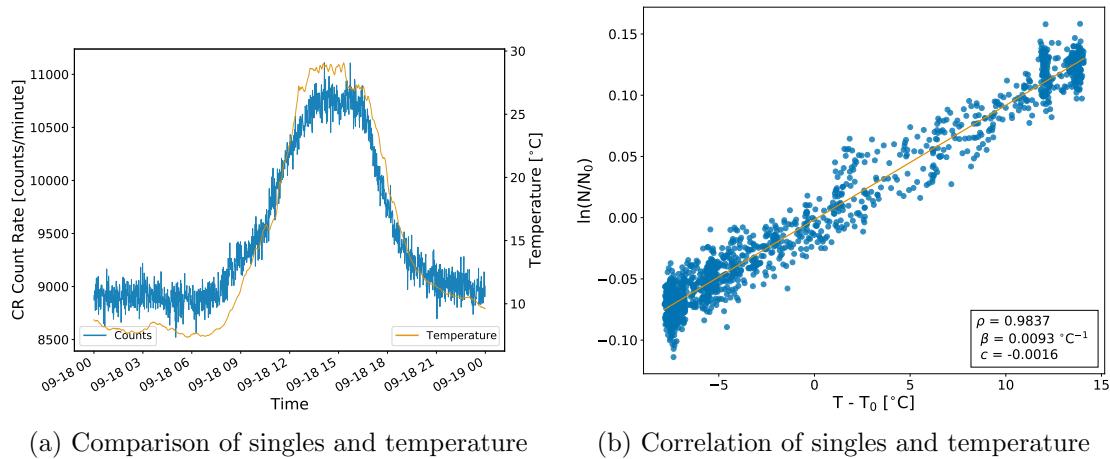


Figure 2.9: The relationship between the singles data and the temperature within the roof box over a single day. (a) Shows the comparison between singles data (blue) and temperature data (orange); (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the 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 1. 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, the raw and corrected singles data are shown in together in Figure 2.10.

In Figure 2.10 we see that the large, diurnal excursions are adequately removed from the singles data after this correction method. This method of temperature

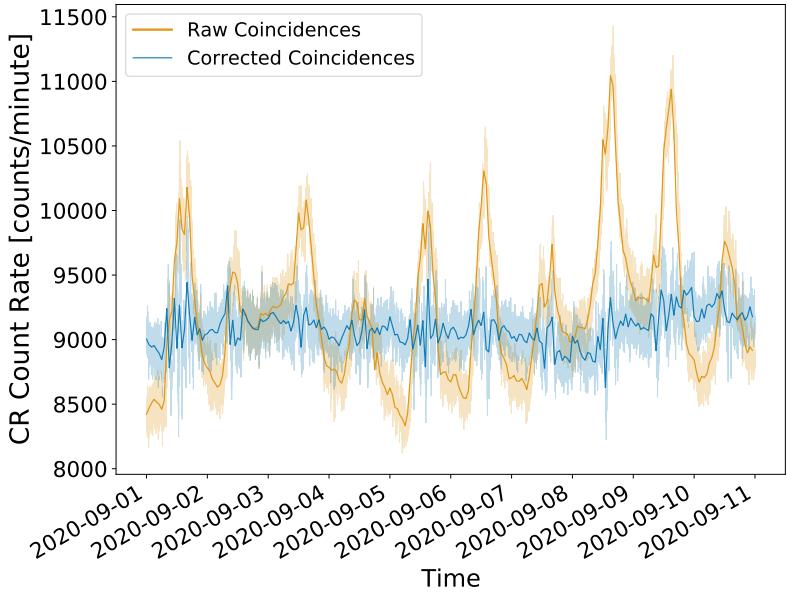


Figure 2.10: The singles data before (orange line) and after (blue line) the temperature correction process. The hourly resampled data are over-plotted to highlight the main variation in the data.

correction was routinely applied to the singles data. However, the corrected data were not stored on the HiSPARC servers.

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 2.11 shows a comparison between the the singles of the two stations, and the temperature within the box of station 14008.

In Figure 2.11 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.

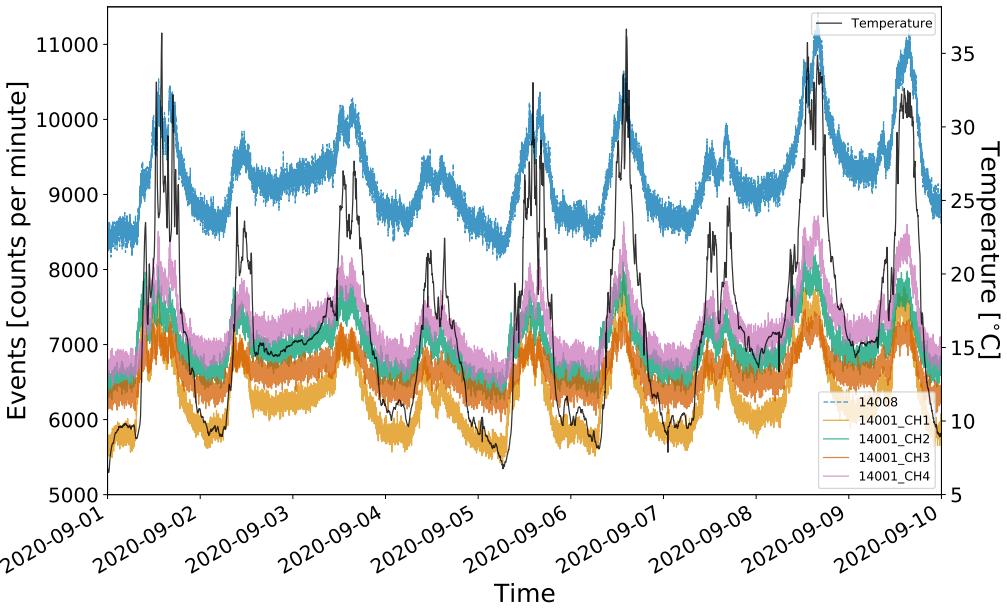


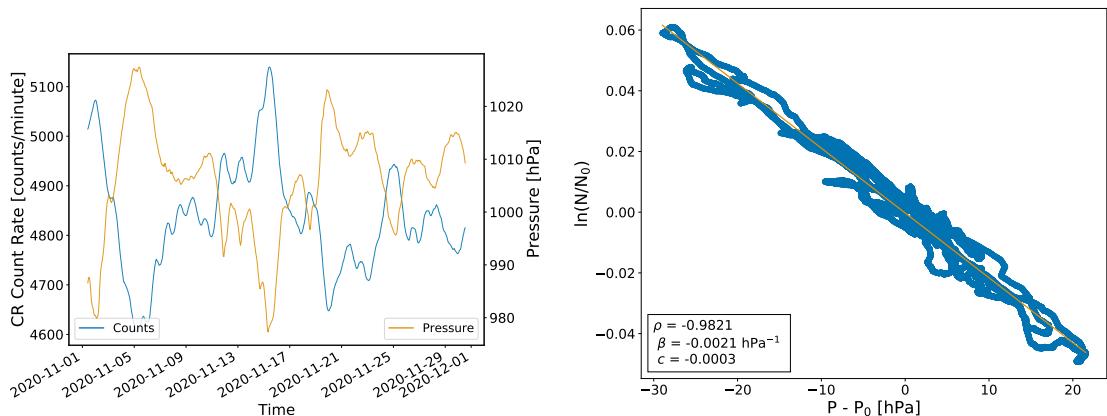
Figure 2.11: Comparison between HiSPARC station 14001 singles data and the singles data and temperature measured by station 14008. Black line: the temperature within the roof box of station 14008; dashed, blue line: HiSPARC station 14008 singles data; solid, orange, green, red, and purple lines: HiSPARC station 14001 singles data.

2.5.2 Barometric Correction

Using the method outlined in Section 2.4 we were able to perform the barometric correction between the interpolated pressure data and coincidences and singles data. Figure 2.12 shows a comparison plot of the smoothed coincidences and atmospheric pressure data sets.

As expected, Figure 2.12 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 2.13. It is clear from Figure 2.13 that the large excursions are adequately removed from the data after the correction.

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



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

Figure 2.12: The relationship between the smoothed, full coincidences data and the atmospheric pressure. (a) Shows the comparison between coincidences data (blue) and pressure data (orange), both with a 12-hour box-bar smoothing applied, to highlight the relationship; (b) shows the correlation between the coincidences counts and pressure, (blue points) and the fitted line (orange) to calculate the correction coefficient.

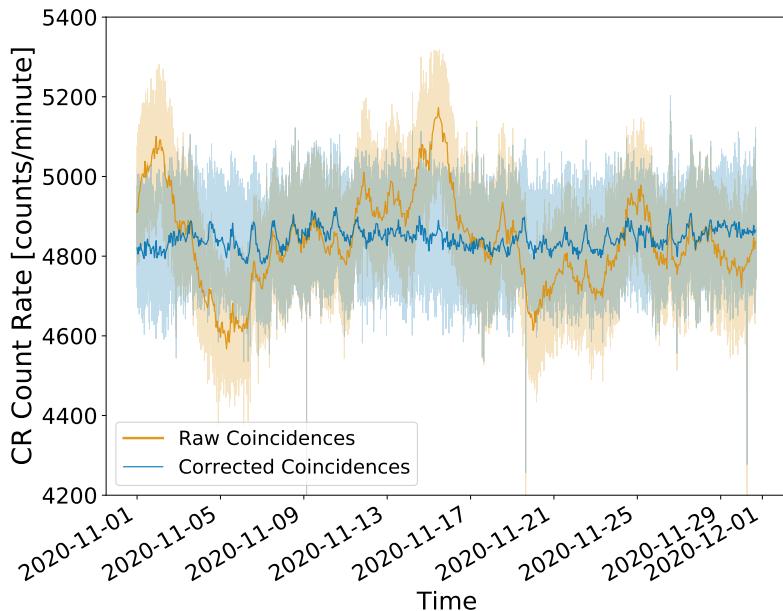


Figure 2.13: Showing the coincidences data before (orange line) and after (blue line) the barometric correction. The hourly resampled data are over-plotted to highlight the main variation in the data.

configuration. However, these data were not stored on the Raspberry Pi or local servers.

2.6 Results

2.6.1 Observations

From the Cosmic Ray Simulations for Kascade (CORSIKA) AS simulations performed in Chapter 1, we predicted a ground level muon rate passing 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., 2020](#); [Particle Data Group et al., 2020](#)).

In Figure 2.14 we show the corrected, full coincidence data for a typical day. 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 HiSPARC events and singles data in Chapter 1. For any given epoch, the diurnal effect can be removed, if necessary, by de-trending the data with a 12- or 24-hour box-car smoothing, which is also shown in Figure 2.14.

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. Convergence was interrogated using the \widehat{R} diagnostic factor using the criteria that chains did not converge if $\widehat{R} > 1.01$.

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 $\sim 80 \mu/\text{s}$ in this stacked detector configuration. This is lower

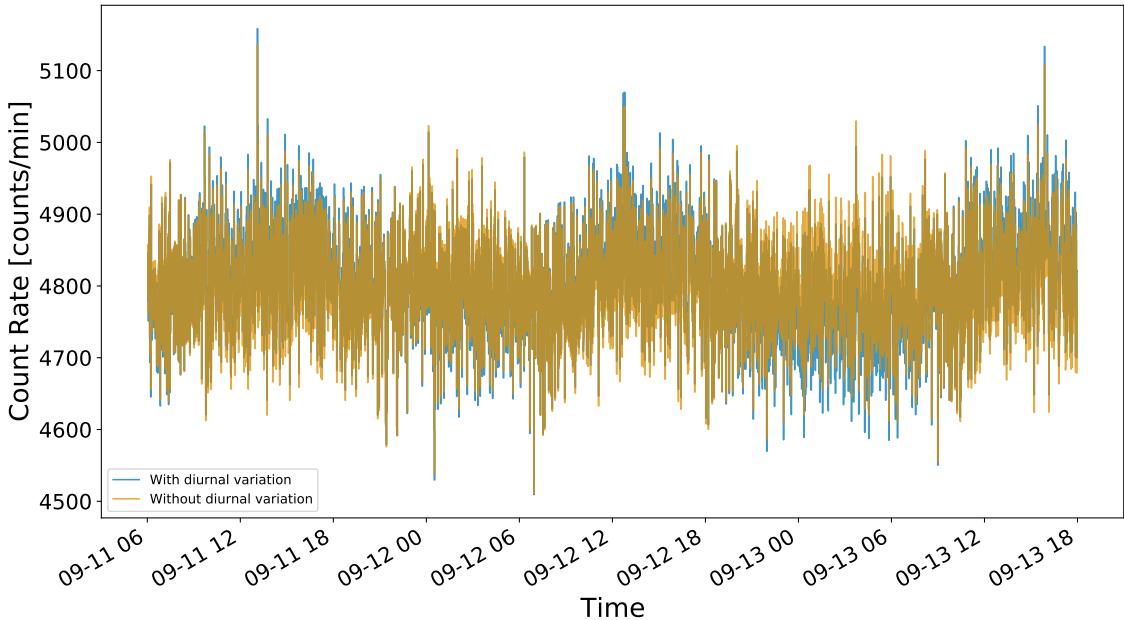


Figure 2.14: Time series of coincidences data, corrected for atmospheric pressure. The blue line shows the corrected data displaying the diurnal variation with a peaks at around midday. The orange line shows the corrected coincidences de-trended using a 12-hour smoothing window, to remove the diurnal effect.

than the estimates predicted in the simulations in Chapter 1, but is close to the predicted values for the non-vertical simulations, which represent a good approximation to the true muon flux at ground level. With a count rate of count rate of $\sim 80 \mu/\text{s}$, the Poisson noise is a rate of $\sim 9 \mu/\text{s}$, which represents $\sim 11\%$ of the signal.

These observations have used the full coincidences data, to determine the mean count rate. These data are stored only locally, but we also acquire the reduced count rates which is 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 are acquired slightly differently. As discussed in Section 2.3.4, the reduced counts (stored locally by the Python script) use 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 2.15 we show a comparison between the reduced coincidences data stored locally and those recorded as events data in the HiSPARC server.

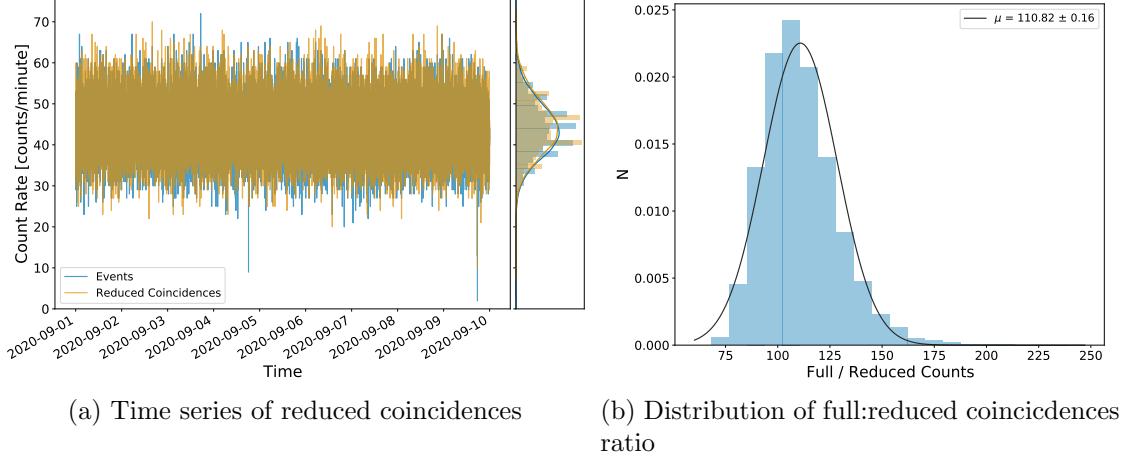


Figure 2.15: (a) Comparison of the reduced coincidences stored locally (orange) and as events data in the HiSPARC server (blue). (b) Distribution of the ratio of the full and reduced coincidence rates.

The mean value of the Poisson distribution of reduced coincidences data is ~ 44 counts/min ($\sim 0.73 \mu\text{s}$), which is a reduction by a factor of ~ 110 from the full coincidences data. We can see from Figure 2.15 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: $1 \mu\text{s}$), coincidence window (duration: $1.5 \mu\text{s}$), and a post-trigger window (duration: $3.5 \mu\text{s}$). The duration of these coincidence windows means that any delay on the order of $\sim 10 \text{ ns}$ is small and does not effect 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 sever are measuring the same signal, with a count rate of count rate of $\sim 0.73 \mu\text{s}$.

To understand the noise properties of this new configuration we investigated the random noise which is induced by random, spurious counts between both PMTs and

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 ~ 120 ns was added between the two signals. The Full Width at Half Maximum (FWHM) of a typical pulse from the PMTs is ~ 25 ns, and the total duration from beginning-to-end is on the order of 100 ns ([van Dam et al., 2020](#)), therefore the ~ 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 are shown in Figure 2.16. We can see that the noise is nominally ~ 1 count/minute.

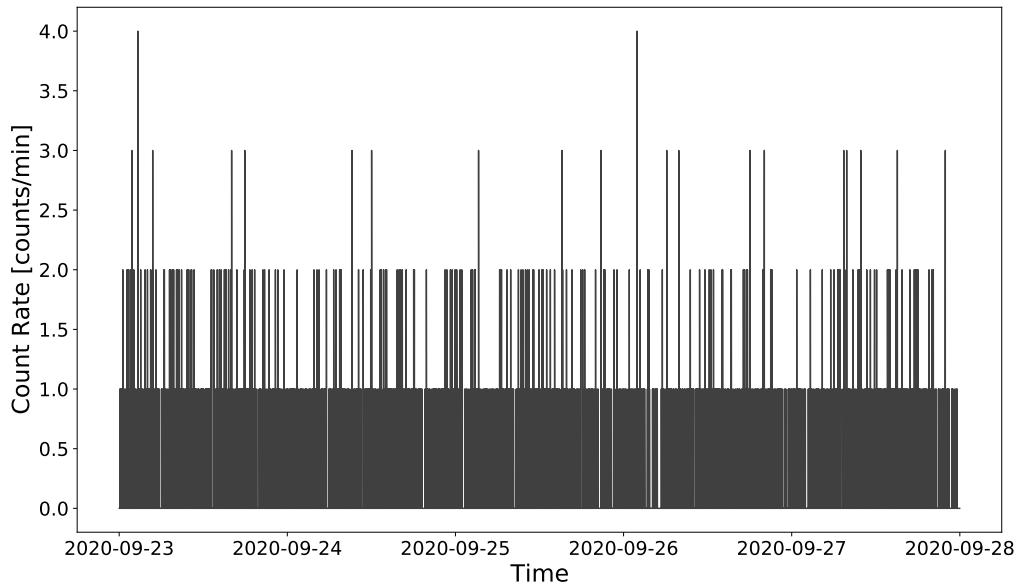


Figure 2.16: Time series of spurious coincidences data measured over five days.

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 \hat{R} diagnostic factor using the criteria that chains did not converge if $\hat{R} > 1.01$. The distribution of the random coincidences is shown in Figure 2.17.

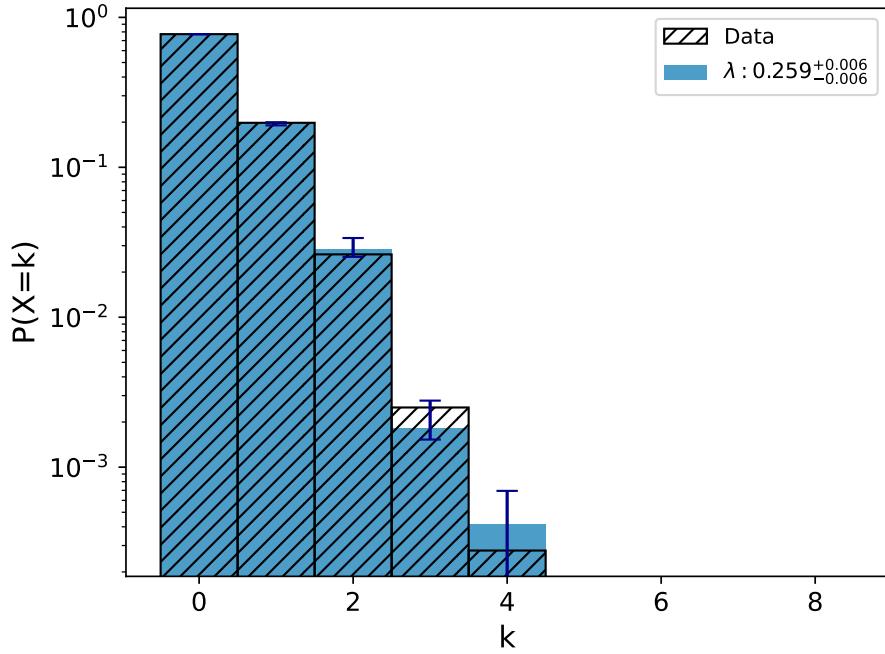


Figure 2.17: Distribution plot of the random coincidences data (hatched bars), and probability mass function of a Poisson distribution with mean equal to the median value of the posterior distribution after sampling (blue bars). Blue error bars represent the 68% credible intervals either side of the median.

The median value of the posterior distribution for 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 small noise; under this Poisson likelihood function the probability of no noise is $\sim 77\%$, 1 count/minute is $\sim 20\%$, and over 2 counts/minute is $< 3\%$.

It is also important to note that in Figure 2.16, there is no diurnal signal 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.

2.6.2 Comparison with Neutron Monitors

It is useful to compare the data from this new HiSPARC station to an existing Neutron Monitor (NM) detector, as this can highlight whether any confirmed space weather events using the Global Neutron Monitor Network (GNMN) have been observed with this new instrument. 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 1 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 ~ 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 2.18, a comparison is shown between the corrected HiSPARC coincidences data and the Dourbes NM station during November 2020.

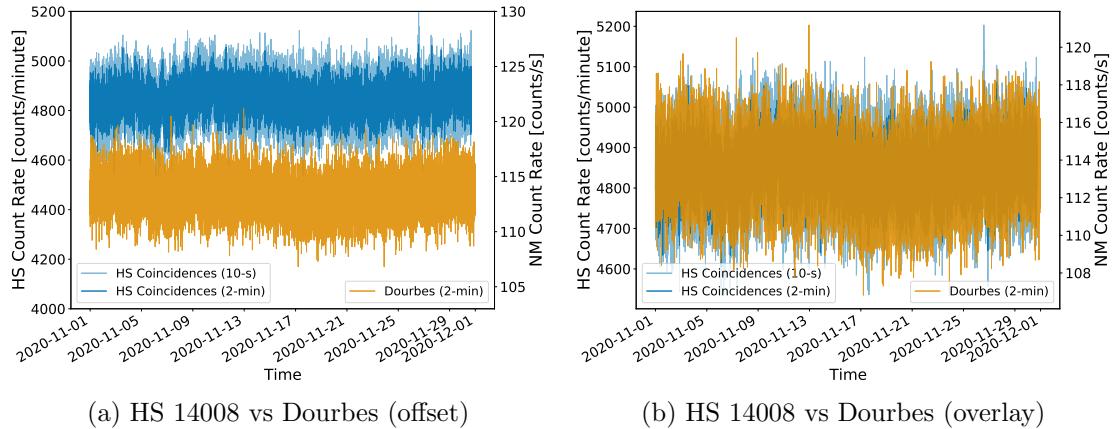


Figure 2.18: A comparison between the corrected HiSPARC station 14008 coincidences data (blue) and the pressure corrected neutron monitor data measured at the Dourbes NM station, Belgium. (a) Shows the two data sets offset and (b) overlayed, showing the similarities between the two data sets.

The plots in Figure 2.18 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 $\rho \sim 0.38$ 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) is 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.

2.6.3 Single Station Space Weather Uses

Simulations of artificial data were performed to determine the magnitude of GLEs that may be observed in this new detector configuration, as described in Section 2.4.2. As we have shown, the HiSPARC 14008 station has a background, mean count rate, $\lambda \sim 80 \mu/\text{s}$, and a noise of $\sigma \sim 0.26 \mu/\text{min}$. These were used as inputs to the simulations, where GLEs were simulated with amplitudes of: 1.0%–5.0%, in steps of 0.5%, and 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 of the resultant data for each amplitude of GLE, compared to the background signal/mean count rate of the detector, without a GLE signal. An example of the output from the statistical tests on a single iteration is shown in Figure 2.19, for a 10% GLE magnitude and a table showing the average number of statistically significant measurements for a simulation without an injected GLE is shown in Table 2.4.

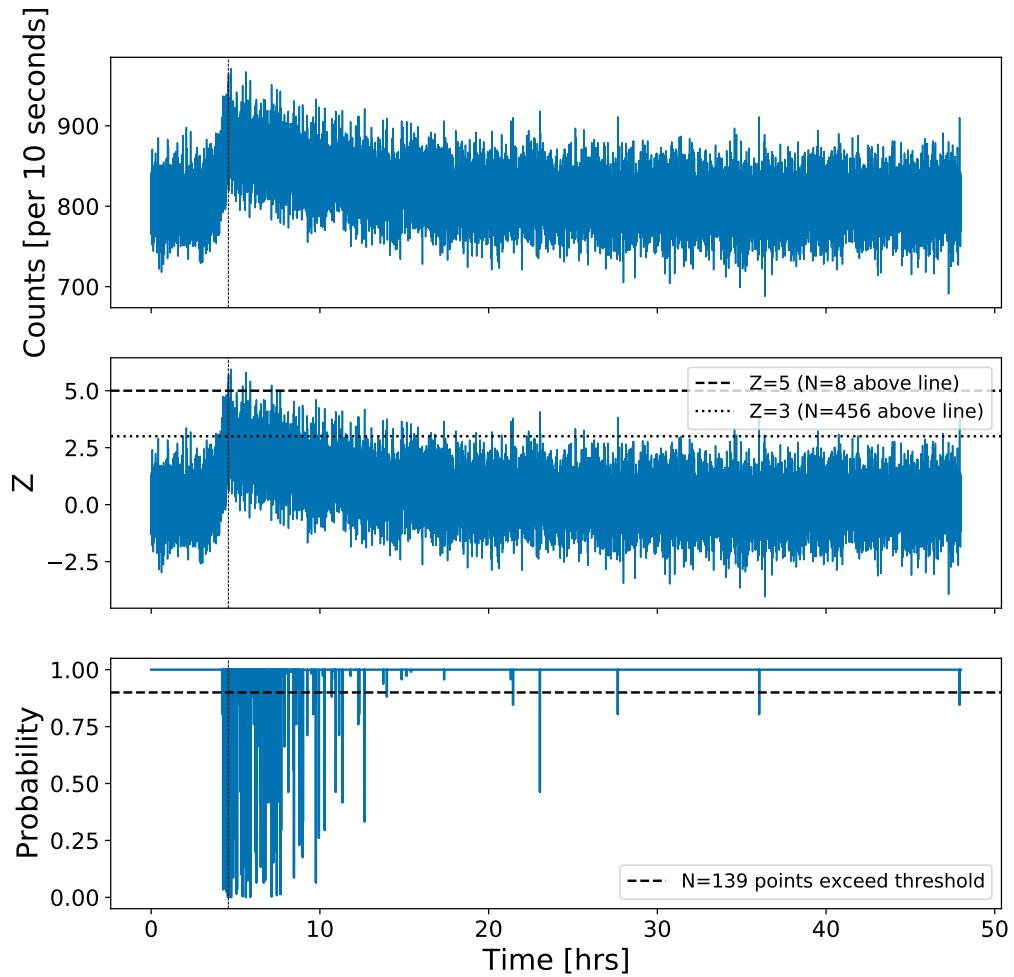


Figure 2.19: Single realisation of a simulation with a 10% GLE magnitude. Top panel shows the raw, artificial data from the simulation, with a 10-second cadence. Middle panel shows the number of measurements exceeding the $Z = 3$ and $Z = 5$ thresholds. The $Z = 3$ and $Z = 5$ thresholds are depicted 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 statistical significance. The dashed line shows the $p = 10\%$ threshold.

In the artificial data, we used the number of statistically significant measurements

to warn whether there existed a possible event in the data, and closer inspection of the statistics tests verified or excluded the existence of a true event. For example, in Figure 2.19, there were 139 points exceeding the 10% binomial limit, and 456 and 8 point above the $Z = 3$ and $Z = 5$ limits, respectively. This indicated the existence of an event in the data. On closer inspection it was clear from the grouping of the number of points exceeding each of the thresholds where the event occurred, at $\sim 4 - 5$ hours.

Table 2.4: Average number of statistically significant measurements in the artificial data with no injected GLEs. The values are for each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for the raw data and averaging over 1-minute and 5-minutes.

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

The results in Table 2.4 provide the expected number of statistically significant fluctuations in the noise for each test method applied to the raw and resampled data. For a particular two day observation, we can say that any epoch with statistically significant peaks in excess of these values can be treated as statistically significant, and likely that there was a space weather event. We expect that any measurements which exceed the $Z = 5$ 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 observing a true event if the number of significant measurements exceeds ~ 33 , ~ 6 , and ~ 2 , for the raw data, 1-minute averaged data, and 5-minute averaged data, respectively. Finally, in the case of the binomial test, the average number of statistically significant fluctuations in the noise exceeding the 10% threshold is in the range $\sim 2 - 4$ in two days of data. We expect that the binomial method is a good measure of the existence of a

true event, due to the consistency of the sensitivity, regardless of averaging over the data.

For each GLE magnitude we ran 1000 iterations of the simulations and were able to average over the number of statistically significant events observed for a given GLE magnitude. This was done for the raw, 10-s cadence data and further simulations were performed for 1-minute averaging and 5-minute averaging, independently.

To summarise the results of all the simulations, Figure 2.20 shows the average number of cadences with statistically significant 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 random significant observations, 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 statistically significant observations for different GLE magnitudes, with error bars representing the 68% credible intervals either side of the median.

We see from Figure 2.20 that in the case where no averaging of data was performed, using the $Z = 3$ significance level, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $> 3.0 - 3.5 \%$. Or using the binomial measure of significance, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $> 4.5 - 5.0 \%$. To be able to see the results clearly, we only show the results for GLE magnitudes of up to 5 %, and in Figure 2.20a, we see using the $Z = 5$ significance level, there are no significant observations for magnitudes $\leq 5 \%$. In the full analysis, we investigated up to 10 %, and determined that at the $Z = 5$ significance level we can differentiate the increase in the number of significant events for GLE magnitude of $\geq 7.5 \%$. These limiting magnitudes are larger than typical magnitudes of GLEs predicted in Section 1.8.4, hence 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.

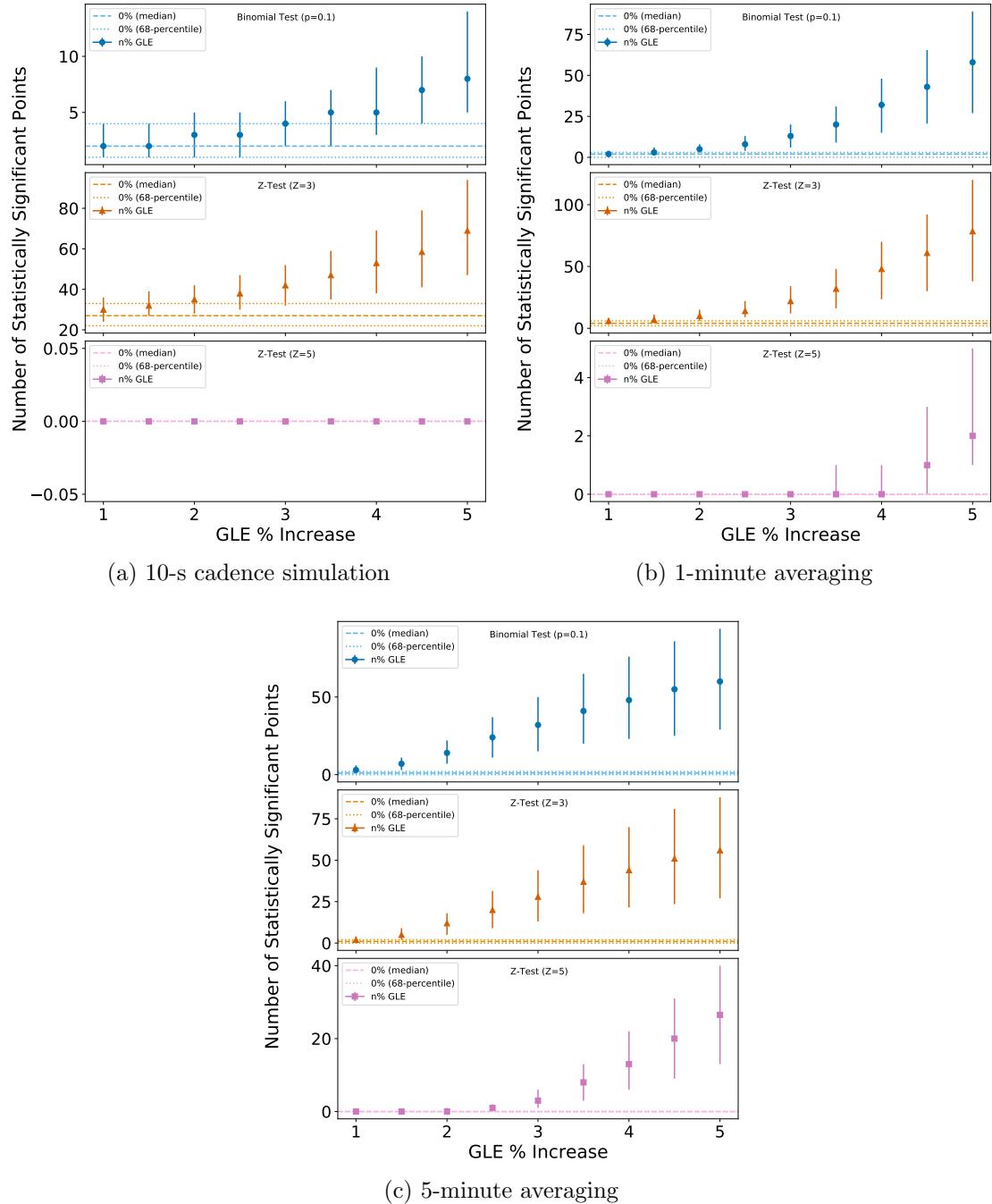


Figure 2.20: Summary plots showing the number of statistically significant measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for the 10-s cadence data; (b) 1-minute averaged data; (c) 5-minute averaged data. 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.

When averaging the data over 1-minute and 5-minute intervals, the detectable GLE magnitude at the $Z = 3$ significance level drops to $2.0 - 2.5\%$ and $\sim 1.5\%$, respectively. Similarly, at the $Z = 5$ significance level this drops to 4.5% and $2.5 - 3.0\%$, respectively. Finally, using the binomial measure of significance at the $p = 10\%$ level, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $2.0 - 2.5\%$ and $\sim 1.5\%$, respectively.

We also note that with an increasing GLE magnitude we observe a larger spread in the average number of statistically significant points. This arises due to the random sampling of GLE properties, i.e. the rise and decay times. A GLE with a longer decay time will result in a higher number of statistically significant measurements, 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 2.20.

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 1.8.4, 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 $> 3\%$ (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).

2.6.4 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 were all geographically close, such that the delay between the signals measured by individual detectors were minimal. This is for instance, assuming 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 statistically significant events observed for a given GLE magnitude. The GLEs were simulated with amplitudes of: 1.0%, 2.0%, 3.0%, 4.0%, and 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 2.5 shows the average number of statistically significant measurements simulations without an injected GLE.

Table 2.5: Average number of statistically significant measurements in the mean of the artificial data for multiple stations, with no injected GLEs. The values are for each of the three statistical tests: binomial, $Z = 3$, and $Z = 5$, for tests with 1, 2, 5, and 10 stations.

No. stations	No. significant measurements		
	Binomial	$Z = 3$	$Z = 5$
1	2^{+2}_{-1}	27^{+6}_{-5}	0 ± 0
2	2^{+3}_{-1}	26^{+22}_{-14}	0 ± 0
5	2^{+4}_{-2}	25^{+23}_{-15}	0 ± 0
10	2^{+3}_{-2}	25^{+32}_{-16}	0 ± 0

To summarise the results of all the simulations, Figure 2.21 shows the average number of statistically significant measurements against the simulated GLE magnitude for the 10-s cadence observations, for 2, 5, and 10 stations. The horizontal, dashed lines shows the median number of random significant observations, i.e. the data in Table 2.5; the horizontal, dotted lines represent the 68% credible intervals either side of the median. Each point shows the median number of statistically significant observations for different GLE magnitudes, with error bars representing the 68% credible intervals either side of the median.

We see from Figure 2.21 that our ability to differentiate the number of significant measurements for a GLE improves with the number of stations used. For a single station, using the binomial 10% threshold, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $> 4.5 - 5.0\%$. Increasing the number of stations to 2, 5, and 10 lowers the observable GLE magnitude to $> 3.0 - 4.0\%$, $\sim 3.0\%$, and $\sim 2.0 - 3.0\%$, respectively.

For a single station above the $Z = 3$ significance level, we expect to be able to differentiate the increase in the number of significant measurements for a GLE with magnitude of $\sim 3.0 - 4.0\%$; however, increasing the number of stations to 2, 5,

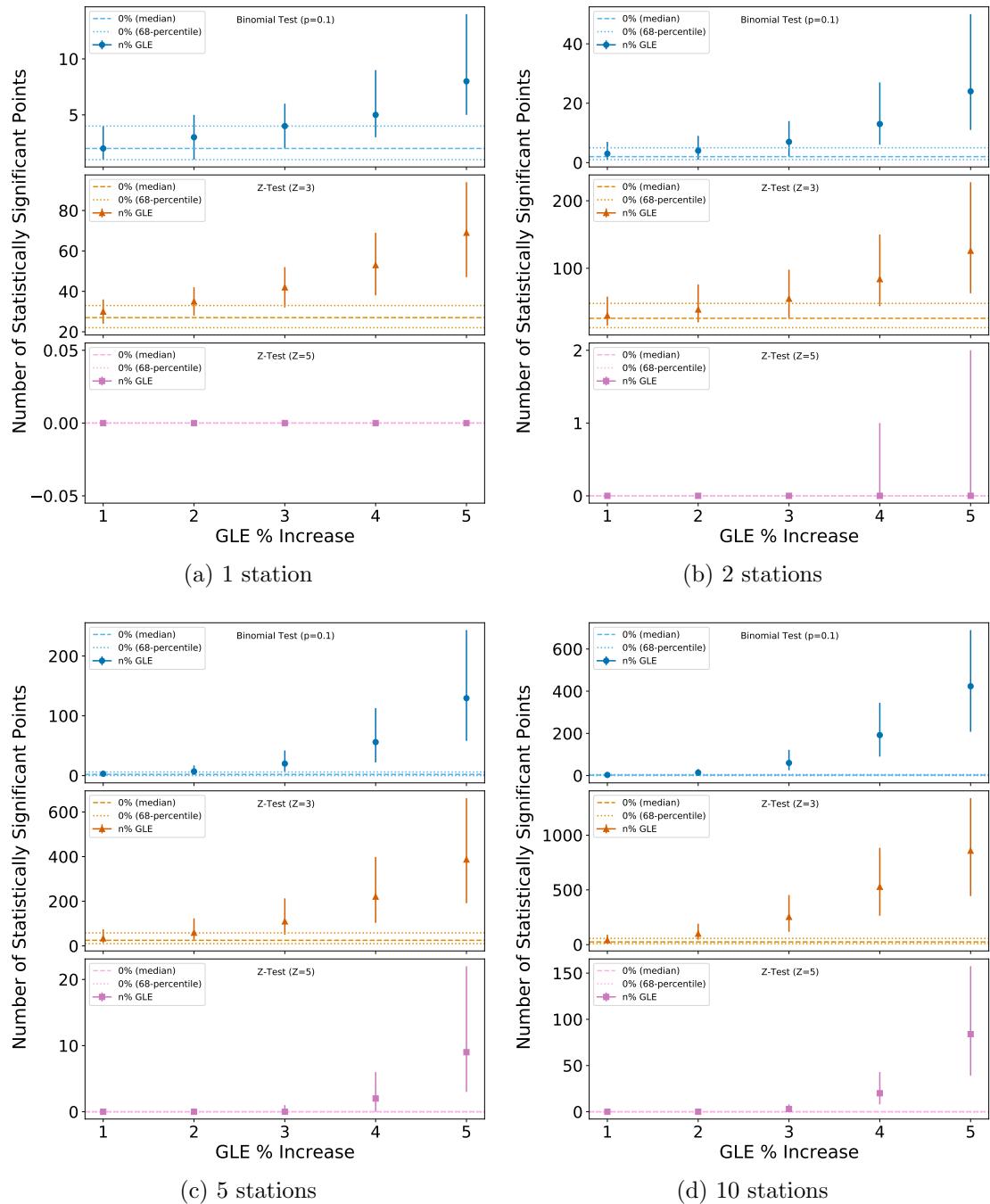


Figure 2.21: Summary plots showing the number of statistically significant measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) Shows the results for a single station; (b) 2 stations; (c) 5 stations; (d) 10 stations . In each plot, the top panel shows the number of 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.

and 10 did not have a significant impact on our ability to observe lower magnitude events. The performance remained similar for 2 and 5 stations, and only lowered to $\sim 2.0 - 3.0\%$ for 10 stations.

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 $\sim 5.0\%$, $\sim 4.0 - 5.0\%$, and $\sim 3.0 - 4.0\%$, respectively.

Despite improving the sensitivity with an increasing number of stations, these limiting magnitudes are larger than typical magnitudes of GLEs predicted in Section 1.8.4. 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 2.6.3, we again observe that increasing GLE magnitude increases the spread in the average number of statistically significant 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 2.21.

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 decay, and a fixed rise time of ~ 30 minutes based on the findings from [Strauss et al. \(2017\)](#). The resultant CCF plots are shown in Figure 2.22, 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.

As the simulated data all experienced the GLE at the same time, there is a clear peak in the CCF at a time shift of zero hours, as expected, showing a strong

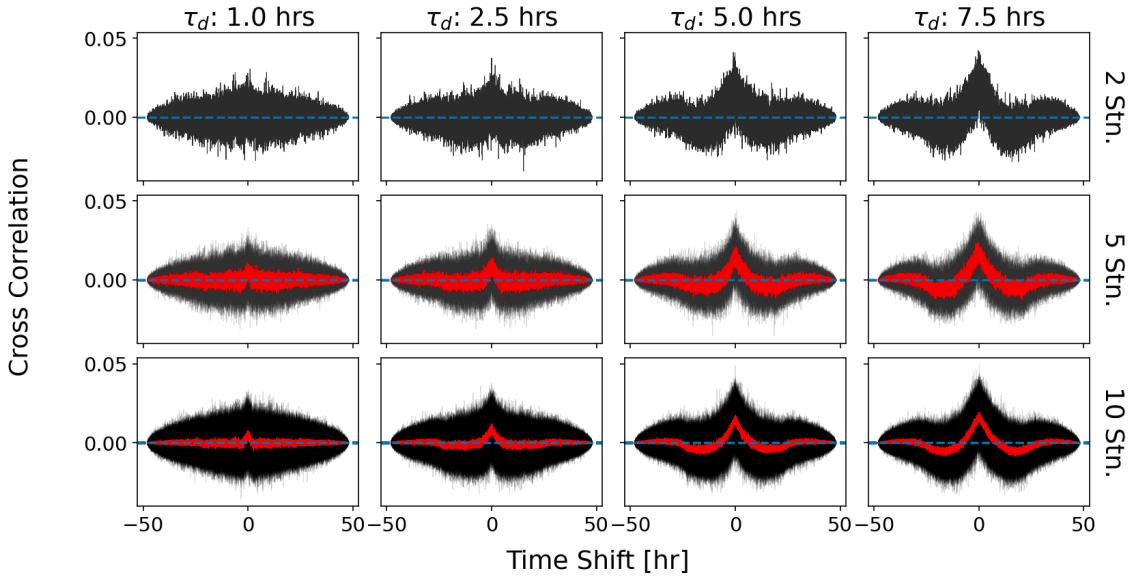


Figure 2.22: Cross-correlation analyses between 2, 5, and 10 stations for a 2% GLE magnitude with varying decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the red line shows the mean of all the realisations. Finally, the dashed, horizontal lines depict a correlation of zero.

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 2.22 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 2.22 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 2.22, 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 2.23.

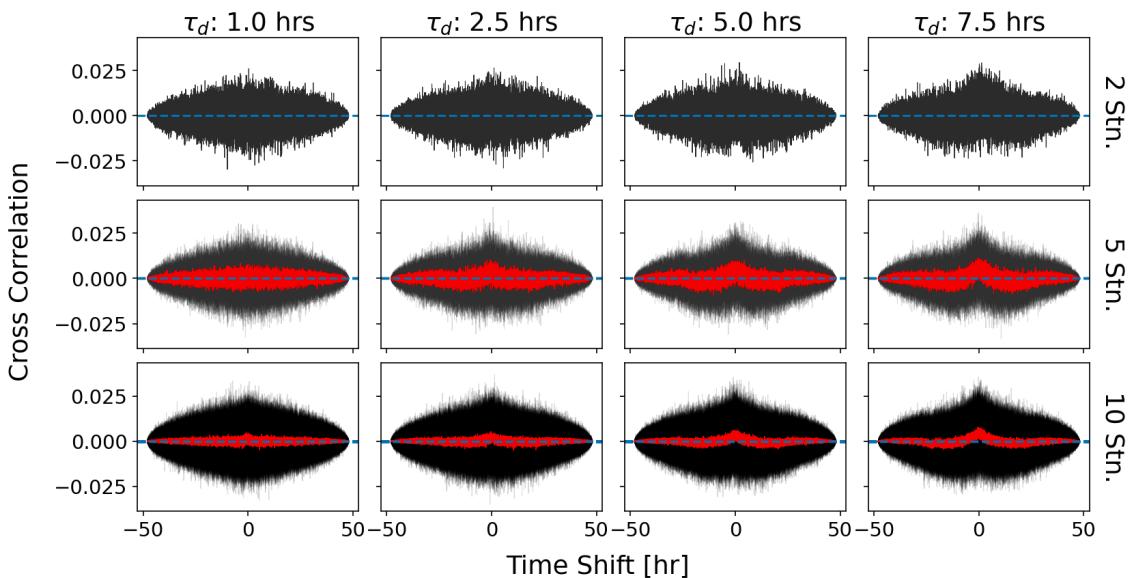


Figure 2.23: Cross-correlation analyses between 2, 5, and 10 stations for a 1% GLE magnitude with varying decay times. The columns show the results for constant decay times, varying the number of stations, and vice versa for the rows. Black lines show the individual realisations of the CCFs, while the 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 2.23, 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 2.23 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 to observe GLEs using the HiSPARC Muon Detectors (MDs) with magnitudes on the order of, or less than, $\sim 1\%$, based on the results shown in Section 1.8.4. In addition, the average decay time of GLEs, as measured by [Strauss et al. \(2017\)](#), is $1.8^{+1.9}_{-1.3}$ hours. 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.

2.7 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 full 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 ~ 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 full 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 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 created which is more relevant for space weather applications as it removes the bias towards higher energy PCRs that the existing HiSPARC network suffers from, and it also reduces the atmosphere induced diurnal effects, which are difficult to correct in the data from the existing HiSPARC stations.
2. The mean count rate in this configuration is $\sim 80 \mu/\text{s}$ and the noise from spurious counts is of about 0.26 ± 0.01 counts/min. This noise is small and infinitesimal 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 statino 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 $> 3\%$ (i.e. similar to that predicted for GLE 42). Through averaging the data into 1- or 5-minute bins, we reduce the noise and

expect to be able to observe GLEs with magnitudes $> 1.5\%$. This is in line with some of the predicted GLE magnitudes from Chapter 1.

5. Through simulating the performance of a network of detectors in this configuration we showed that we can increase the sensitivity of this configuration to observe GLEs on the order of, or less than, $\sim 1\%$, but note that there is a strong dependence on decay time. 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.

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 simplistic 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:

$$y(t) = \sigma + \lambda(1+A) \cdot \begin{cases} \exp\left[\frac{t-t_0}{\tau_r}\right] & \text{if } t \leq t_0 \\ \exp\left[-\frac{t-t_0}{\tau_d}\right] & \text{if } t > t_0 \end{cases}, \quad (\text{B.1})$$

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.

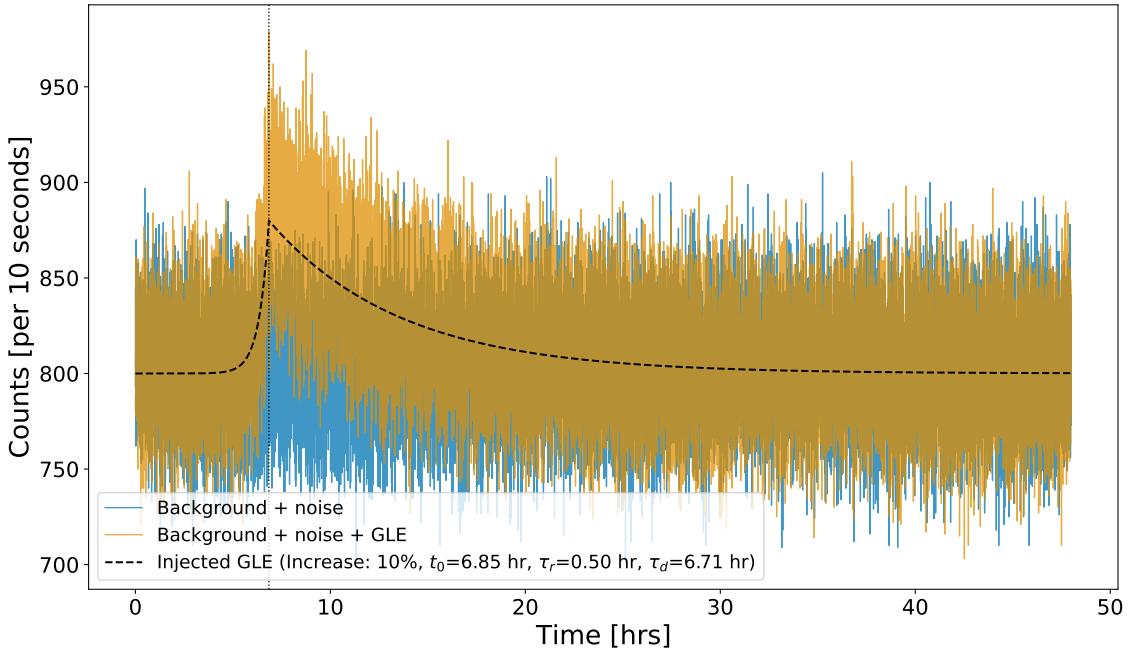


Figure B.1: Example of a realisation of the simulations, to generate an artificial time series with a 10% GLE injected. The black line shows the model for the GLE. The blue line shows the realisation of the signal without a GLE injected, and the orange line shows the realisation of the model with the 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.

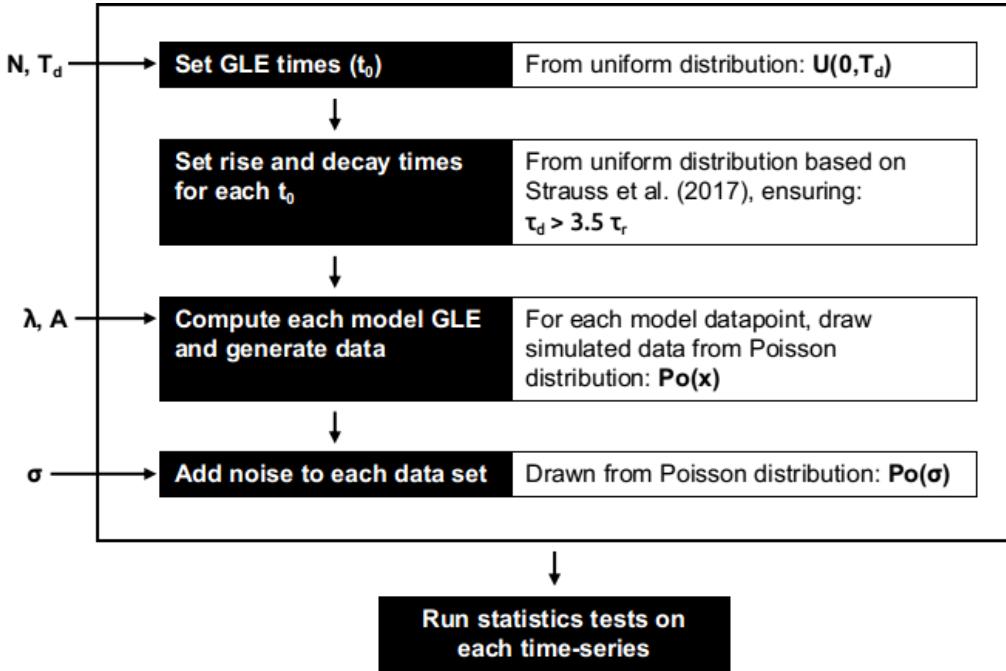


Figure B.2: Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.

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 $\tau_d > 3.5 \tau_r$, (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

- Arunbabu K. P., et al., 2015, *Astronomy and Astrophysics*, 580, A41
- 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
- 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
- Basu S., Chaplin W. J., 2017, Asteroseismic Data Analysis: Foundations and Techniques
- 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
- Berkova M. D., Belov A. V., Eroshenko E. A., Yanke V. G., 2011, *Bull. Russ. Acad. Sci. Phys.*, 75, 820
- Blackmore E. W., Stukel M., Trinczek M., Slayman C., Wen S., Wong R., 2015, *IEEE Transactions on Nuclear Science*, 62, 2792
- Braun I., Engler J., Hrandel J. R., Milke J., 2009, *Advances in Space Research*, 43, 480
- 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
- Cecchini S., Sioli M., 2000, arXiv:hep-ex/0002052
- Corti C., Potgieter M. S., Bindi V., Consolandi C., Light C., Palermo M., Popkow A., 2019, *ApJ*, 871, 253

- Cramp J. L., 1996, PhD Thesis, University of Tasmania, Hobart, <https://core.ac.uk/download/pdf/33327719.pdf>
- Danilova O. A., Demina I. M., Ptitsyna N. G., Tyasto M. I., 2019, *Geomagnetism and Aeronomy*, 59, 147
- 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., Flckiger 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., 2004, in Astrophysics and Space Science Library, Cosmic Rays in the Earths 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., 2010, *Phys.-Usp.*, 53, 496
- Dubey A., Kumar S., Dubey S. K., 2016, *Journal of Physics: Conference Series*, 755, 012056
- Dumbović M., Vršnak B., Čalogović J., 2016, *Sol Phys*, 291, 285
- Dyer C. S., Lei F., Clucas S. N., Smart D. F., Shea M. A., 2003, *Advances in Space Research*, 32, 81
- 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
- Groom D. E., Mokhov N. V., Striganov S. I., 2001, *Atomic Data and Nuclear Data Tables*, 78, 183
- 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

- Humble J. E., Duldig M. L., Smart D. F., Shea M. A., 2012, *Geophysical Research Letters*, 18, 737
- Kudela K., Storini M., Hofer M. Y., Belov A., 2000, *Space Science Reviews*, 93, 153
- 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
- LeCroy 1996, LeCroy Research Systems Model 365AL, Model 465, and Technical Data: Model 622 Logic Units, <https://teledynelecroy.com/lrs/dsheets/365al.htm>
- Lei F., Clucas S., Dyer C., Truscott P., 2004, *IEEE Transactions on Nuclear Science*, 51, 3442
- 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>
- McCracken K. G., Moraal H., Shea M. A., 2012, *ApJ*, 761, 101
- 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
- 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
- Montanus H., 2017, PhD thesis, University of Amsterdam, https://www.nikhef.nl/pub/services/biblio/theses_pdf/
- NMDB 2018, NMDB Event Search Tool (NEST), <http://www.nmdb.eu/nest/>
- 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)
- Ostapchenko S., 2006, *Nuclear Physics B - Proceedings Supplements*, 151, 143
- Owens M. J., Forsyth R. J., 2013, *Living Rev. Sol. Phys.*, 10, 5
- 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., Koutroumpia I., 2013, *New Astronomy*, 19, 10

- Pereira E. G., Vasconcellos C. A. Z., Hadjimichef D., 2020, arXiv:2011.06480 [hep-ex, physics:hep-ph]
- Pesnell W. D., 2020, *Journal of Space Weather and Space Climate*, 10, 60
- Poluianov S., Usoskin I. G., 2017, in Proceedings of Science. Bexco, Busan, Korea, <https://pos.sissa.it/301/#session-3145>
- Rockenbach M., et al., 2014, *Space Sci Rev*, 182, 1
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Comput. Sci.*, 2, e55
- Savić M., Maletić D., Joković D., Veselinović N., Banjanac R., Udovičić V., Dragić A., 2015, *J. Phys.: Conf. Ser.*, 632, 012059
- Schwenn R., 2006, *Living Reviews in Solar Physics*, 3, 2
- 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
- Strauss R. D., Ogunjobi O., Moraal H., McCracken K. G., Caballero-Lopez R. A., 2017, *Solar Physics*, 292, 51
- Thébault E., et al., 2015, *Earth, Planets and Space*, 67, 79
- Thomas S., Owens M., Lockwood M., Owen C., 2017, *Ann. Geophys.*, 35, 825
- Timashkov D. A., et al., 2008, *Astroparticle Physics*, 30, 117
- Tsyganenko N. A., 1989, *Planetary and Space Science*, 37, 5
- Tsyganenko N. A., 2013, *Annales Geophysicae*, 31, 1745
- 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>
- Wawrzynczak A., Alania M. V., 2010, *Advances in Space Research*, 45, 622
- eFoam 2017, SF38 High Foam Datasheet, <https://www.efoam.co.uk/docs/SF38.pdf>
- 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., 2020, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 959, 163577