

# 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

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>List of Abbreviations</b>	<b>vi</b>
<b>1 HiSPARC Station 14008</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Aims . . . . .	2
1.3 HiSPARC 14008 Detector Set-up . . . . .	2
1.3.1 Configuration . . . . .	2
1.3.2 Calibration . . . . .	6
1.3.3 Monitoring Temperature . . . . .	7
1.3.4 Data Acquisition . . . . .	8
1.3.5 Monitoring Pressure . . . . .	11
1.4 Methodology . . . . .	11
1.4.1 Atmospheric Corrections . . . . .	11
1.4.2 Observations . . . . .	12
1.5 Atmospheric Corrections . . . . .	15
1.5.1 Barometric Correction . . . . .	15
1.5.2 Temperature Correction . . . . .	16
1.6 Results . . . . .	17
1.6.1 Observations . . . . .	17
1.6.2 Comparison with Neutron Monitors . . . . .	22
1.6.3 Single Station Space Weather Uses . . . . .	24
1.6.4 Multiple Station Space Weather Uses . . . . .	29
1.7 Conclusions . . . . .	32
<b>A Simulations of the Ground Level Enhancements for the Analysis of HiSPARC Station 14008</b>	<b>34</b>
A.1 Model . . . . .	34
A.2 Configuration of the Simulations . . . . .	35
<b>Bibliography</b>	<b>37</b>

# List of Figures

1.1	Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box/ski box. . . . .	3
1.2	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. . . .	3
1.3	Schematic diagram of the HiSPARC 14008 station NIM crate configuration. 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 and this shows half the signal interfaces directly with the HiSPARC electronics box and half the signal is passed through the NIM crate for processing. . . . .	5
1.4	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). . .	7
1.5	(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) the sensor within the roof boxes, located by the PMTs. The temperature sensor is soldered into the circuit board in the top-middle of the image. . . . .	8
1.6	Schematic diagram of the HiSPARC station 14008 data acquisition interfaces. . . . .	9
1.7	The relationship between the ... (a) shows the ... and pressure data; (b) shows the correlation between the singles counts and pressure, and the fitted line to calculate the correction coefficient. . . . .	15
1.8	Showing the coincidence data before (orange line) and after(blue line) the barometric correction. . . . .	16
1.9	... . . . .	16
1.10	The relationship between the ... (a) shows the ... and temperature data; (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the correction coefficient. . .	17
1.11	Time series of coincidences data, corrected for atmospheric pressure and displaying the diurnal variation with a peak at around midday. .	18

1.12	(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. . . . .	19
1.13	Time series of random coincidences data... . . . .	21
1.14	Distribution of random coincidences data... and Poisson distribution of the random coincidences, along with the median posterior fitted mean of the sample... . . . .	22
1.15	... . . . .	23
1.16	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. . . . .	25
1.17	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. 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. . . . .	27
1.18	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) ...; (b) ...; (c) ...; (d) ... . 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. 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. . . . .	30
1.19	... for 2% GLE magnitude ... . . . .	31
1.20	... for 1% GLE magnitude ... . . . .	32

A.1	...	35
A.2	Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.	36

# List of Tables

1.1	Variables stored in the coincidences files of the HiSPARC 14008 instrument. . . . .	9
1.2	Variables stored in the temperature files of the HiSPARC 14008 instrument. . . . .	10
1.3	... with no injected GLEs ... . . . .	24
1.4	... with no injected GLEs ... . . . .	29

# List of Abbreviations

**CCF** Cross-Correlation Function.

**CEDA** Centre for Environmental Data Analysis.

**CORSIKA** Cosmic Ray Simulations for Kascade.

**CR** Cosmic Ray.

**FWHM** Full Width at Half Maximum.

**GLE** Ground Level Enhancement.

**GNMN** Global Neutron Monitor Network.

**HiSPARC** High School Project on Astrophysics and Research with Cosmics.

**HMC** Hamiltonian Monte Carlo.

**HV** High Voltage.

**MIDAS** Met Office Integrated Data Archive System.

**NERC** Natural Environment Research Council.

**NIM** Nuclear Instrumentation Module.

**NM** Neutron Monitor.

**NUTS** No U-Turn Sampler.

**PCR** Primary Cosmic Ray.

**PMT** Photo Multiplier Tube.

**SCR** Solar Cosmic Ray.

**SEP** Solar Energetic Particle.

**STFC** Science and Technology Facilities Council.

**TTL** Transistor-Transistor Logic.

# 1 HiSPARC Station 14008

## 1.1 Introduction

It was shown in Chapter ??, using data from the High School Project on Astrophysics and Research with Cosmics (HiSPARC) network, that it was not clearly capable of observing space weather events and this is also hindered as they are rather sensitive to variations in the terrestrial conditions.

To some extent, it was possible to eliminate the variation in Cosmic Rays (CRs) due to terrestrial variation from the HiSPARC data; however it was shown to be not always so simple, as different detectors in the HiSPARC network showed different responses to pressure and temperature variation. The non-linear relationship between temperature and CR count means the correction of the count rate due to thermal fluctuations is non-trivial, unlike the counterpart correction for pressure.

It is believed that the atmospheric thermal fluctuations 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 sky boxes are in direct sunlight.

An instance of thermal noise in a single PMT will be random, and uncorrelated with an instance of thermal noise in another PMT. It is therefore possible to hypothesise that it is unlikely that within the coincidence window of  $\sim 1.5 \mu\text{s}$ , that a coincidence between 2 PMTs would be due to random thermal noise induced in the PMTs.



To exploit this, it is possible to stack 2 detectors on top of each other to measure a single muon which traverses both scintillators, hence inducing signals in both PMTs.

## **1.2 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 standardise a configuration that allows for the observation of space weather events.

We aimed to set up a new detector, perform 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 properties to understand what magnitude of GLE could be observed with the new set-up. This would help us to understand how likely it was to observe the any space weather events with the alternative HiSPARC configuration.

## **1.3 HiSPARC 14008 Detector Set-up**

### **1.3.1 Configuration**

The configuration of HiSPARC station 14008 is shown in Figure 1.1; 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. Previously, we could only count single muons in the singles rates, but we have shown that the data is inconsistent between stations and it is difficult to disentangle the effect of temperature, PMT noise, and the diurnal effect. This design limits the energy bias in the events data, which is biased to higher energy Primary Cosmic Rays (PCRs) because of the required footprint 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 removing thermal fluctuations.

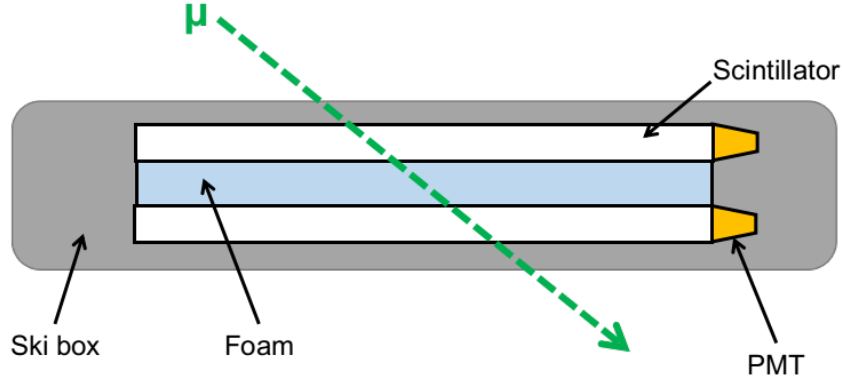


Figure 1.1: Schematic diagram of the HiSPARC station 14008 detector set-up within the roof box/ski box.

To protect the scintillators and PMTs, we sandwiched a layer of high density ( $\rho = 38 - 40 \text{ kg m}^{-3}$ , [eFoam \(2017\)](#)) foam, of thickness,  $\Delta x = 50 \text{ mm}$ , between the scintillators, as can be seen on the lab work bench in Figure 1.2a. Upon the completed assembly of the detectors, they are placed within the roof box on the roof of the Poynting Physics building on the campus of University of Birmingham, as shown in Figure 1.2b.



(a) Scintillators inside roof box



(b) Complete detector on the roof

Figure 1.2: 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.

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:

$$\Delta E = \Delta x S \rho \cos(\theta), \quad (1.1)$$

where  $\Delta x$  is the thickness of the material,  $S$  is the stopping power of the material,  $\rho$  is the density of the material, and  $\theta$  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,  $\rho = 1.03$  g cm<sup>-3</sup> (Montanus, 2017). The stopping power of the scintillator for a minimum ionising particle is  $S \sim 2$  MeV g<sup>-1</sup> cm<sup>2</sup> (Fokkema, 2012; Montanus, 2017). The energy loss of a vertically incident muon in a single detector is therefore  $\Delta E \sim 4$  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, the muons will lose an additional  $\sim 0.4$  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  $\sim 7.42$  MeV.

The standard HiSPARC station set-up is such that the PMTs are connected to the HiSPARC electronics box for data acquisition. In the standard configuration, the trigger rate of events is  $\sim 1$  Hz. In this stacked configuration the trigger rate is significantly higher,  $\sim 70$  Hz; hence the data produced is the equivalent of approximately half of the existing HiSPARC network. The network 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 stack detectors. To acquire the data in this set-up, we used a Nuclear Instrumentation Module (NIM) crate, as shown in Figure 1.3.

The data acquisition is discussed in Section 1.3.4, but here we discuss the configuration of the NIM crate set-up. The High Voltage (HV) signal from the detector

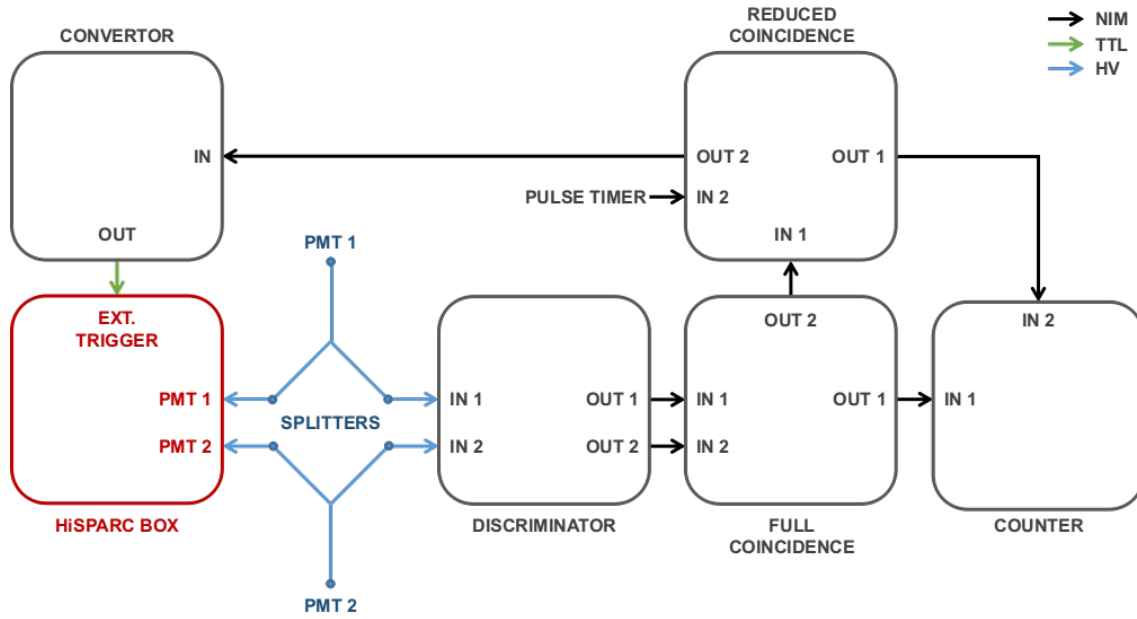


Figure 1.3: Schematic diagram of the HiSPARC 14008 station NIM crate configuration. Black arrows depict a NIM signal; green arrows show a TTL signal; blue arrows depict a HV signal. The PMTs from the detectors are split and this shows half the signal interfaces directly with the HiSPARC electronics box and half the signal is passed through the NIM crate for processing.

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 is first passed 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 signal is reduced in amplitude by a factor of 2; we used a discriminator threshold of  $-35$  mV, i.e. half of the HiSPARC high threshold.

The discriminator outputs two NIM signals which are connected to the first coincidence module (LeCroy model 622). This records every coincidence between the two PMTs and the output from this module is directed to a NIM quad counter/timer module (ORTEC model 974). One channel of the NIM counter records the full coincidences 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  $\sim 1\%$  duty cycle (gate width =  $45.0\,\mu\text{s}$  and repeat 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  $\sim 100$ . 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 acquires the counts directly from the PMTs.

[discussion, and also probably use a table, to discuss the delays introduced by the NIM crate...]

### 1.3.2 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  $700\,\text{V}$  ([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 1.4, 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  $100\text{--}130\,\text{Hz}$  for signal above the high trigger threshold, and singles rate of  $<400\,\text{Hz}$  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 1.4 [UPDATE THIS PLOT...!!!!]. The voltage calibration

plot shows drastically the different performances one can get from different PMTs, therefore it is necessary to treat each PMT individually when calibrating.

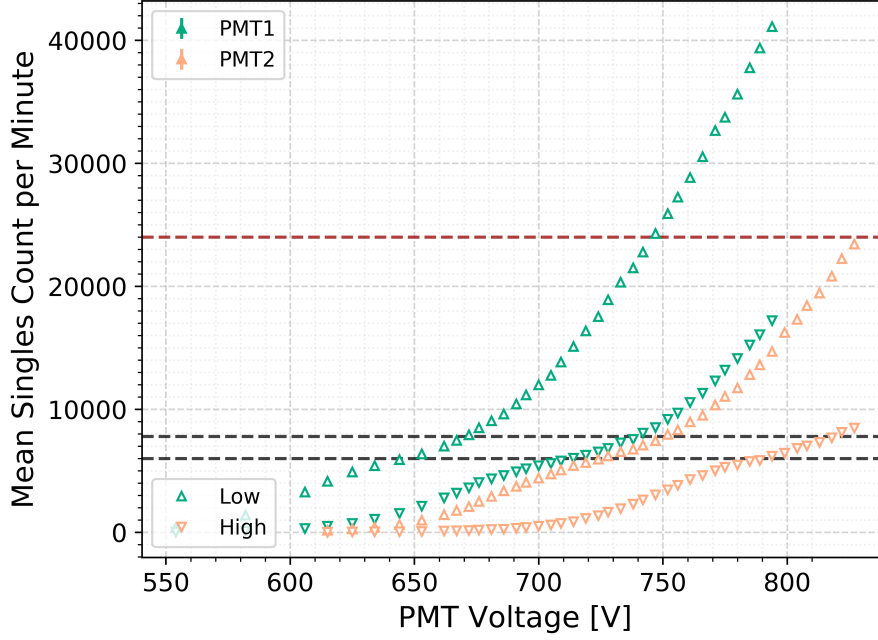


Figure 1.4: Voltage calibration curve for the PMTs of station 14008. The upper, red-dashed line indicates the upper limit for the low threshold singles rate (400 Hz), and the lower 2, black-dashed lines indicate the upper and lower bounds for the high threshold singles rate (100–130 Hz).

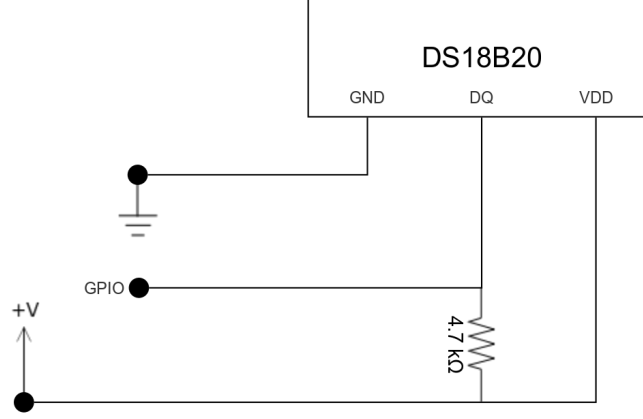
### 1.3.3 Monitoring Temperature

In Chapter ??, we suspected that the singles count rates (and thus event count rates also) were affected by the temperature of the PMT within the HiSPARC roof-boxes.

Some of the existing HiSPARC stations monitor local temperature however none measure the temperature of the PMT within the roof box; therefore the temperature of the PMT itself is unknown, and thus we cannot account for the thermal noise. When building this new HiSPARC station, a temperature sensor was placed into the roof box which allowed us to monitor the temperature.

Figure 1.5a shows the schematic 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 1.3.4). Three wires were used for

the operation of the DS18B20: constant current voltage, ground, and data. The temperature is read on a 10-second cadence and is recorded in degrees Celsius with a precision of  $0.001^{\circ}\text{C}$ .



(a) Circuit diagram



(b) Sensor within roof box

Figure 1.5: (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) the sensor within the roof boxes, located by the PMTs. The temperature sensor is soldered into the circuit board in the top-middle of the image.

### 1.3.4 Data Acquisition

The new HiSPARC station uses two methods of data acquisition. The singles data and reduced coincidences data 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. A schematic diagram showing the interfaces between the Raspberry Pi and the other hardware is shown in Figure 1.6.

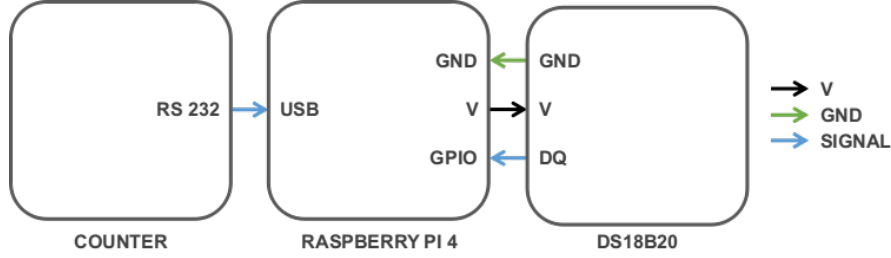


Figure 1.6: Schematic diagram of the HiSPARC station 14008 data acquisition interfaces.

The Raspberry Pi 4 was used to control the data acquisition by running a Python script. 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 second cadence.

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 1.1.

Table 1.1: 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 file are outlined in Table 1.2.

Table 1.2: 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 data is stored locally, but it is also stored on the University of Birmingham particle physics servers as a back-up<sup>1</sup>. Access to the data is not necessarily open, and to request access, one should contact the System Administrator for the Particle Physics Group Computing Facilities.

The reduced coincidences data are acquired using the HiSPARC data acquisition software and are stored within the HiSPARC servers. The HiSPARC servers record this data as the events and the data can be acquired using the methods described in Section ??.

[What is the width of the signals generated by the NIM crate? Is it more or less than the approx. 25ns FWHM of the pulses..? Then relate that to: "The pulse width,  $T_w$ , is important only insofar as it determines the maximum rate of pulses that may be represented by the pulse train, since pulses which occur more frequently than  $1/T_w$  cannot be resolved" ...]

[a few statements on the availability of the data... i.e. it's generally ok and a complete set beyond Feb 2020, but many interruptions due to the pandemic... After

---

<sup>1</sup>Disk location: /disk/moose/general/epesv001/datadisk/147.188.46.117\_hisparc\_pi/

around Sept 2020 the data becomes more reliable...]

### 1.3.5 Monitoring Pressure

As with the previous chapter, it was still necessary to account for the barometric effect on the muon count rate, for both the singles data and the full and reduced coincidences data. To monitor the pressure, a nearby 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 pressure is measured at the MIDAS station level and a correction for altitude is not applied. The station is located in Coleshill, Warwickshire (ID: 19187), nearby Birmingham International Airport,  $\sim 20$  km from the HiSPARC detectors; we believe the pressure variation is small over this distance.

The pressure data is 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.

## 1.4 Methodology

### 1.4.1 Atmospheric Corrections

Using the theory outlined in Section ?? 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 (??). ...

After considerate review of the method of temperature correction, the method also discussed in Chapter ?? was used, i.e. the linear relationship; however, the method was tweaked slightly. The steps for correcting for the effect of temperature

were:

- To remove the long-term temperature trend over a month of data, the CR and temperature data were smoothed using a 24-hr moving mean. Using the linear relationship defined by equation (??), 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 here, 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 ensures that there is a smooth transition in the corrected data between days.

### 1.4.2 Observations

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

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (1.2)$$

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

The Poisson distribution is also additive such that if two variables,  $n_1$  and  $n_2$ , follow Poisson distributions, with mean values  $\lambda_1$  and  $\lambda_2$ , respectively, then the sum also follows a Poisson distribution:

$$P(n; \lambda_1, \lambda_2) = P(n; \lambda_1 + \lambda_2), \quad (1.3)$$

where  $n = n_1 + n_2$ .

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 was created, to simulate the detector's response to space weather events.

Artificial data was generated using the method discussed in Appendix A

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:

$$F(k; \lambda) = \sum_{i=0}^k \frac{\lambda^i e^{-\lambda}}{i!}, \quad (1.4)$$

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

$$p(k) = 1 - F(k; \lambda). \quad (1.5)$$

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  $\lambda$  in  $N$ -cadences is:

$$p_N = 1 - [1 - p(\lambda)]^N, \quad (1.6)$$

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. The probability of finding at least  $r$  occurrences in  $N$ -cadences at or above  $k$ , given the mean level,  $\lambda$ , is given by equation (1.7), which is equal to equation (1.6) when  $r = 1$ ,

$$p[r; p(k), N] = \sum_{r=r}^N \binom{N}{r} p(k)^r [1 - p(k)]^{N-r}. \quad (1.7)$$

By applying equation (1.7) 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 uses 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:

$$s = k - \lambda, \quad (1.8)$$

where  $s$  is the excess in the signal,  $k$  is the measured signal and  $\lambda$  is the expected value, or background signal. The significance can then be approximated by:

$$Z = \frac{s}{\sigma}, \quad (1.9)$$

where  $\sigma$  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

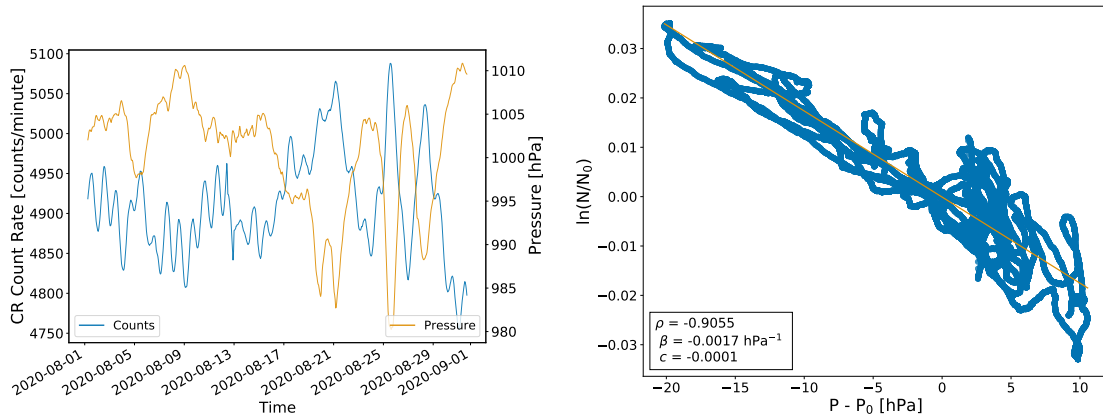
such to allow us to observe lower amplitude space weather events. We also ran the statistics tests on 1-minute and 5-minutes averages of the artificial data. The statistical tests were run under the same underlying principals, however instead of using the Poisson cumulative distribution in equation (1.5), we instead used the Gaussian cumulative distribution for the averaged data with mean,  $\mu = \lambda$ , and standard deviation,  $\sigma = \sqrt{\mu/n}$ , where  $n$  represents the number of data points used in the average. Similarly, using the Gaussian distribution with mean,  $\mu = \lambda$ , and standard deviation,  $\sigma = \sqrt{\mu/n}$ , in equation (1.9).

## 1.5 Atmospheric Corrections

### 1.5.1 Barometric Correction

Using the theory outlined in Section ?? we were able to perform the barometric correction between the 1-minute interpolated pressure data and 1-minute averaged coincidences and singles data.

[show the plot of the correction over a month...]



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

Figure 1.7: The relationship between the ... (a) shows the ... and pressure data; (b) shows the correlation between the singles counts and pressure, and the fitted line to calculate the correction coefficient.

[show a before vs after plot too...]

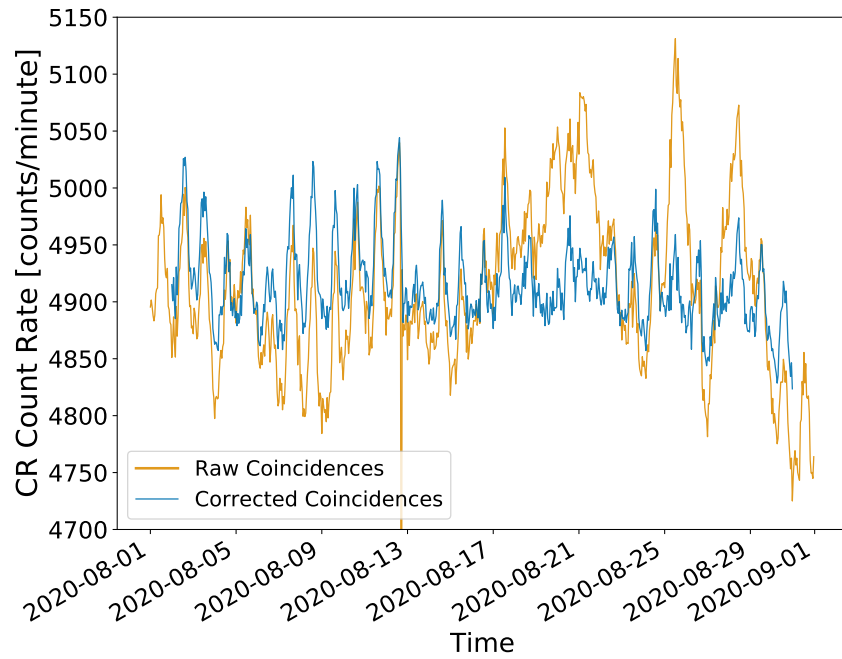


Figure 1.8: Showing the coincidence data before (orange line) and after (blue line) the barometric correction.

## 1.5.2 Temperature Correction

One can see the relationship between the singles rates and the temperature in Figure 1.9...

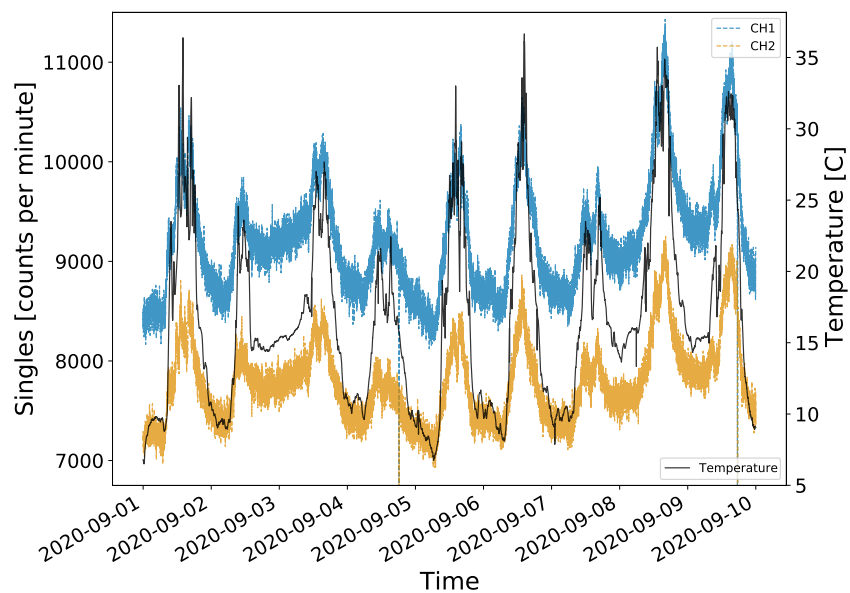


Figure 1.9: ...

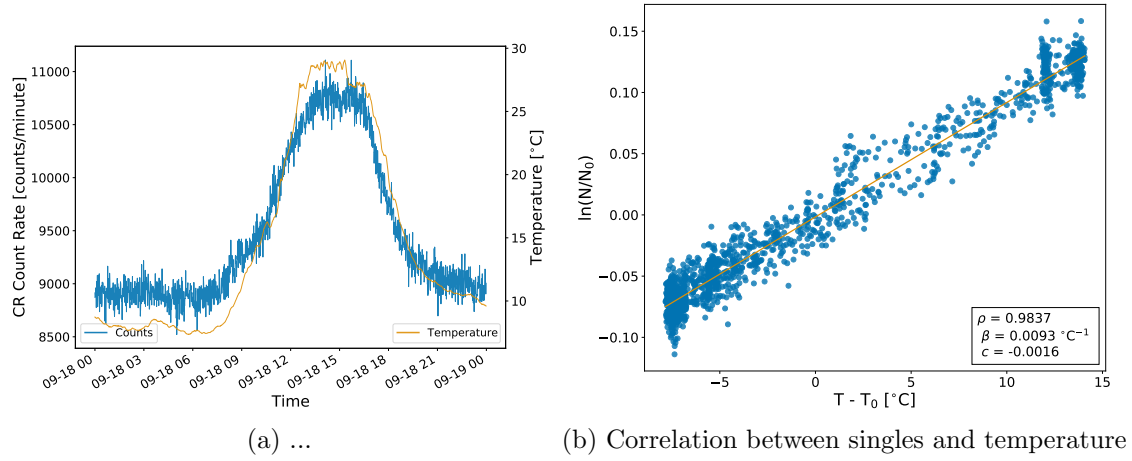


Figure 1.10: The relationship between the ... (a) shows the ... and temperature data; (b) shows the correlation between the singles counts and temperature, and the fitted line to calculate the correction coefficient.

An additional benefit of the temperature monitor at station 14008 is that it is suitable for providing an estimate of the temperature inside the roof-boxes of stations 14001, which is located on the same roof.... [show figure of the singles for both stations and the temperature!]

[Note: temperature correction doesn't work/not necessary on the coincidences but it is necessary on the singles rates!]

## 1.6 Results

### 1.6.1 Observations

From the Cosmic Ray Simulations for Kascade (CORSIKA) simulations in Chapter ??, we predicted a ground level muon rate passing through a single HiSPARC detector of  $\sim 85 \mu\text{/s}$  (for non-vertical, i.e.  $70^\circ$  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 1.11 we show the full coincidence count 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  $\sim 1 - 2\%$ , 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 ???. For any given epoch, the diurnal effect can be removed, if necessary, by removing the trend of a 12-hour smoothed data set, which is also shown in Figure 1.11.

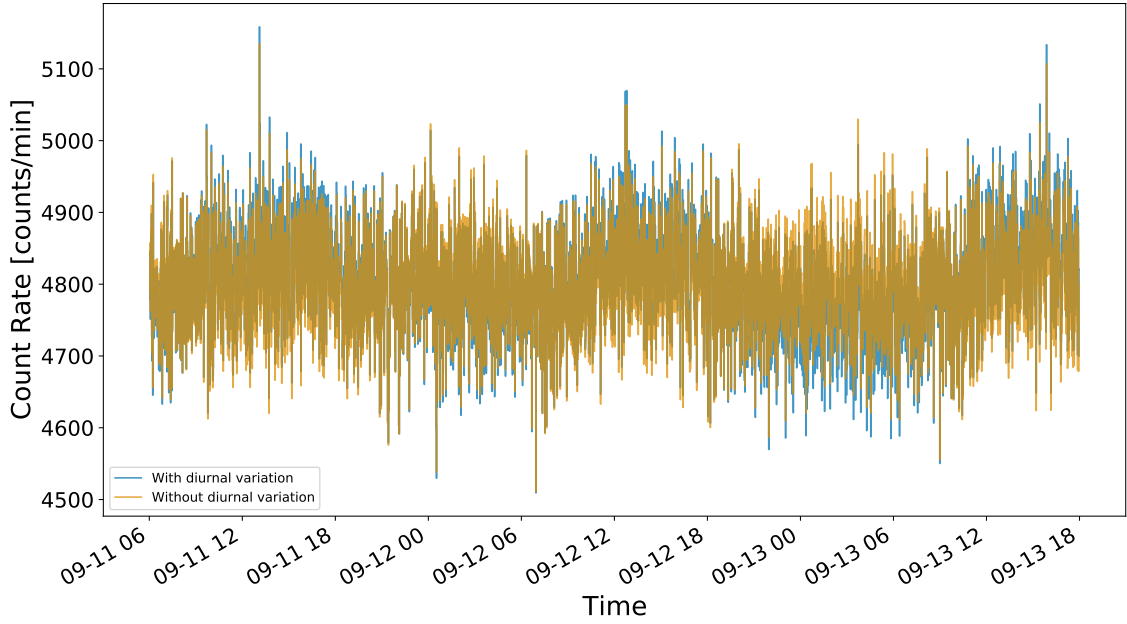


Figure 1.11: Time series of coincidences data, corrected for atmospheric pressure and displaying the diurnal variation with a peak at around midday.

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  $\hat{R}$  diagnostic factor using the criteria that chains did not converge if  $\hat{R} > 1.01$ .

The median value of the posterior distribution for the mean value of the Poisson distribution of these coincidence data is  $4797 \pm 2$  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 than the estimates predicted in the simulations in Chapter ??, 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 shown the full coincidences data, which are stored only locally. The data sent to the HiSPARC servers are the reduced counts, which use the NIM gate signal to reduce the count rate by a factor of  $\sim 100$ . This data is also stored locally, but the data are acquired slightly differently. As discussed in Section 1.3.4, the reduced counts stored locally use the NIM counter to count 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 uses this trigger to read the events directly from the PMTs. Due to the delays in the signal in the NIM create, one could reasonably have a concern that the two sources of data may differ. In Figure 1.12 we show a comparison between the reduced coincidences data stored locally and those recorded as events data in the HiSPARC server.

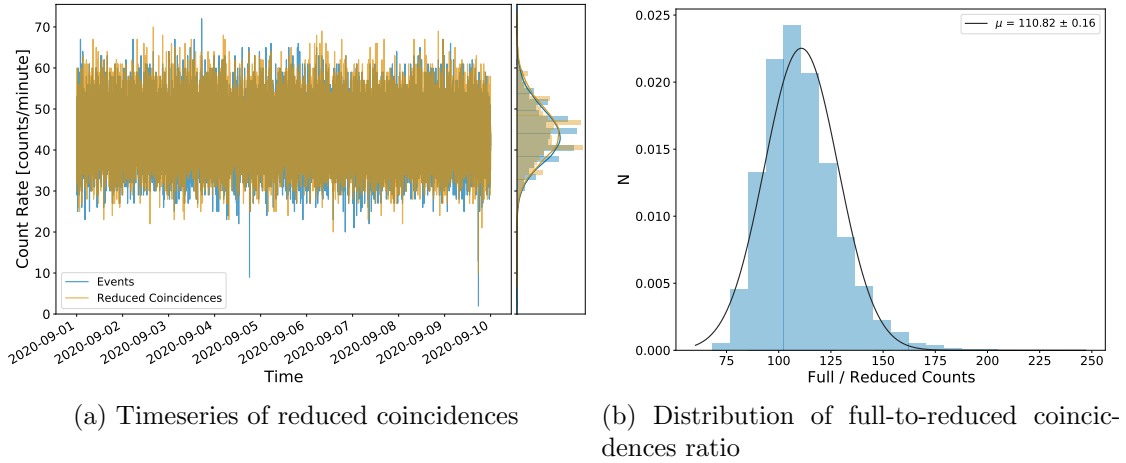


Figure 1.12: (a) Comparison of the reduced coincidences stored locally (orange) and as events data in the HiSPARC server (blue). (b) Distribution of the ratio of the full and reduced coincidence rates.

The mean value of the Poisson distribution of these coincidence data is  $\sim 44 \text{ counts/min}$

( $\sim 0.73 \mu\text{s}$ ), which is a reduction by a factor of  $\sim 110$  from the full coincidences data. We can see from Figure 1.12 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  $\sim 0.73 \mu\text{s}$ , the Poisson noise is on the same order of magnitude as the signal.

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  $\sim 120 \text{ ns}$  was added between the two signals. The Full Width at Half Maximum (FWHM) of a typical pulse from the PMTs is  $\sim 25 \text{ ns}$ , and the total duration from beginning-to-end is on the order of  $100 \text{ ns}$  ([van Dam et al., 2020](#)), therefore the  $\sim 120 \text{ ns}$  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 1.13. We can see that the noise is nominally  $\sim 1 \text{ count/minute}$ .

We know the noise must follow a Poisson distribution, therefore we aimed to quantify the mean value of the noise. The noise was sampled to determine the mean of the Poisson distribution using the `pymc3` NUTS extension to a HMC sampling

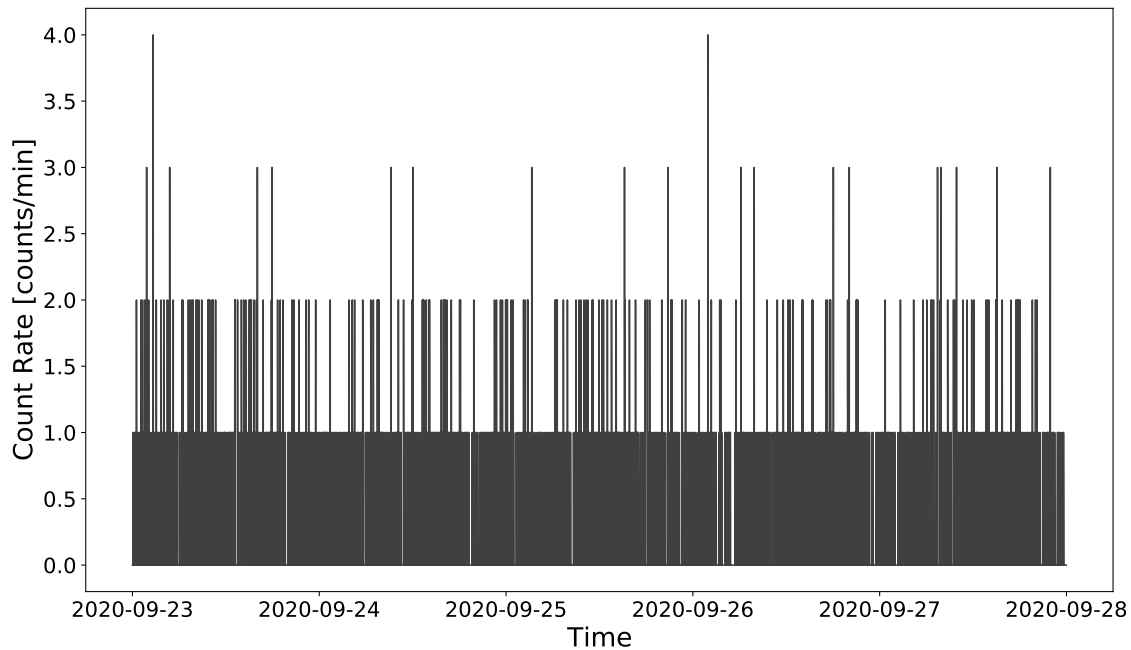


Figure 1.13: Time series of random coincidences data...

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 1.14

The median value of the posterior distribution for the mean value of the Poisson distribution of random coincidence is  $0.259 \pm 0.006$  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 1.13, there is no diurnal signal in the random noise. 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.

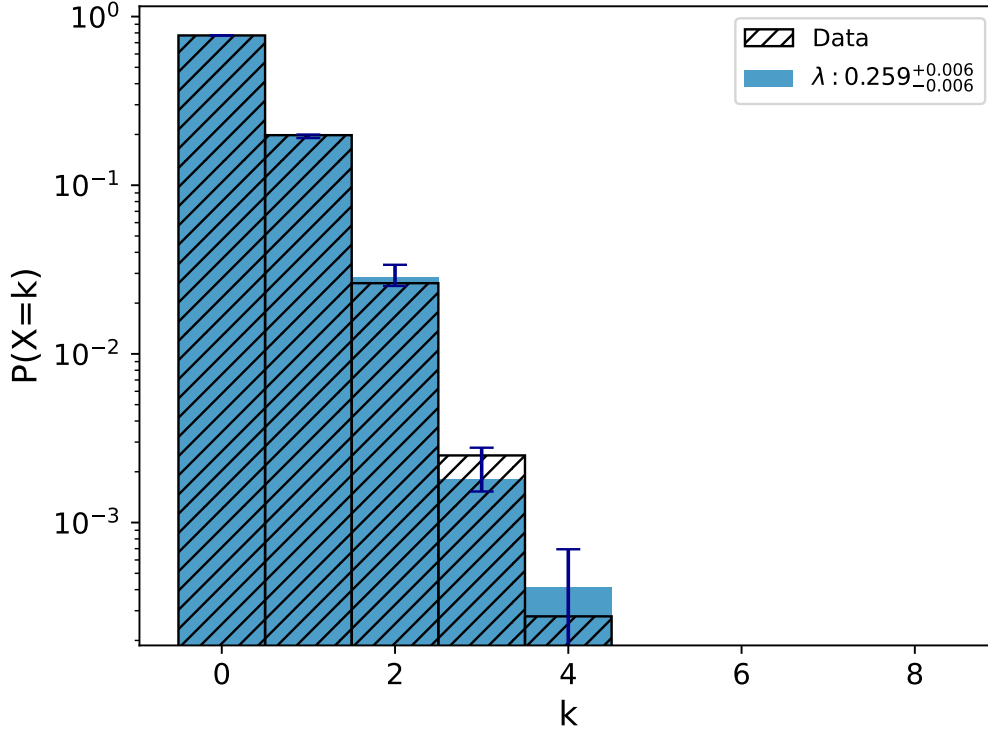


Figure 1.14: Distribution of random coincidences data... and Poisson distribution of the random coincidences, along with the median posterior fitted mean of the sample...

## 1.6.2 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 configuration. 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 ?? had suffered difficulties with data consistency during this epoch, therefore another station was used for the analysis here. We chose to analyse the NM count rate at, Dourbes, which is located in Belgium and is the nearest NM to the HiSPARC network, at  $\sim 525$  km away from stations 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 1.15, a comparison is shown between the HiSPARC data and the Dourbes NM station during November 2020.

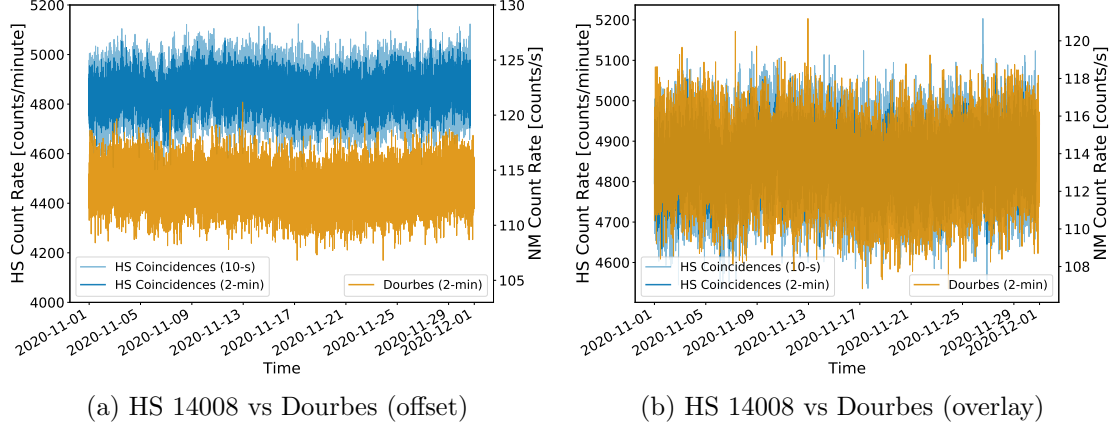


Figure 1.15: ...

The plots in Figure 1.15 show a good visual agreement between the two detectors. Despite a good visual agreement, the Pearson correlation coefficient is  $\rho \sim 0.08$ , highlighting that there is no correlation between the two stations, at the 2-minute level. When resampling the data to 1-hour and 1-day the correlation increases to  $\rho \sim 0.38$  and  $\rho \sim 0.41$ , respectively. This shows a weak, low frequency correlation between the two stations. This weak correlation shows 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 is 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 they are the closest stations to the HiSPARC 14008 detector in Birmingham, it is useful to continue using the Dourbes NM station, but also to incorporate the Kiel NM station, where 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.

### 1.6.3 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 1.4.2. As we have shown, the HiSPARC 14008 station has a background, mean count rate,  $\lambda \sim 80 \mu/\text{s}$ , and a noise of  $\sigma \sim 0.26 \mu/\text{s}$ . These were used as inputs to the simulations, where GLEs were simulated with amplitudes of: 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, 5.0%, 7.5%, and 10.0%. The start times, rise and decay times were all randomly sampled 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 and compare it to the background signal of the detector, without a GLE signal. An example of the statistical tests on a single iteration is shown in Figure 1.16, 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 1.3.

Table 1.3: ... with no injected GLEs ...

<b>Data cadence</b>	<b>No. significant measurements</b>		
	<b>Binomial</b>	<b>Z = 3</b>	<b>Z = 5</b>
10-second	$2_{-1}^{+2}$	$27_{-5}^{+6}$	$0 \pm 0$
1-minute	$2_{-2}^{+1}$	$4 \pm 2$	$0 \pm 0$
5-minute	$1 \pm 1$	$1 \pm 1$	$0 \pm 0$

The results in Table 1.3 provides 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

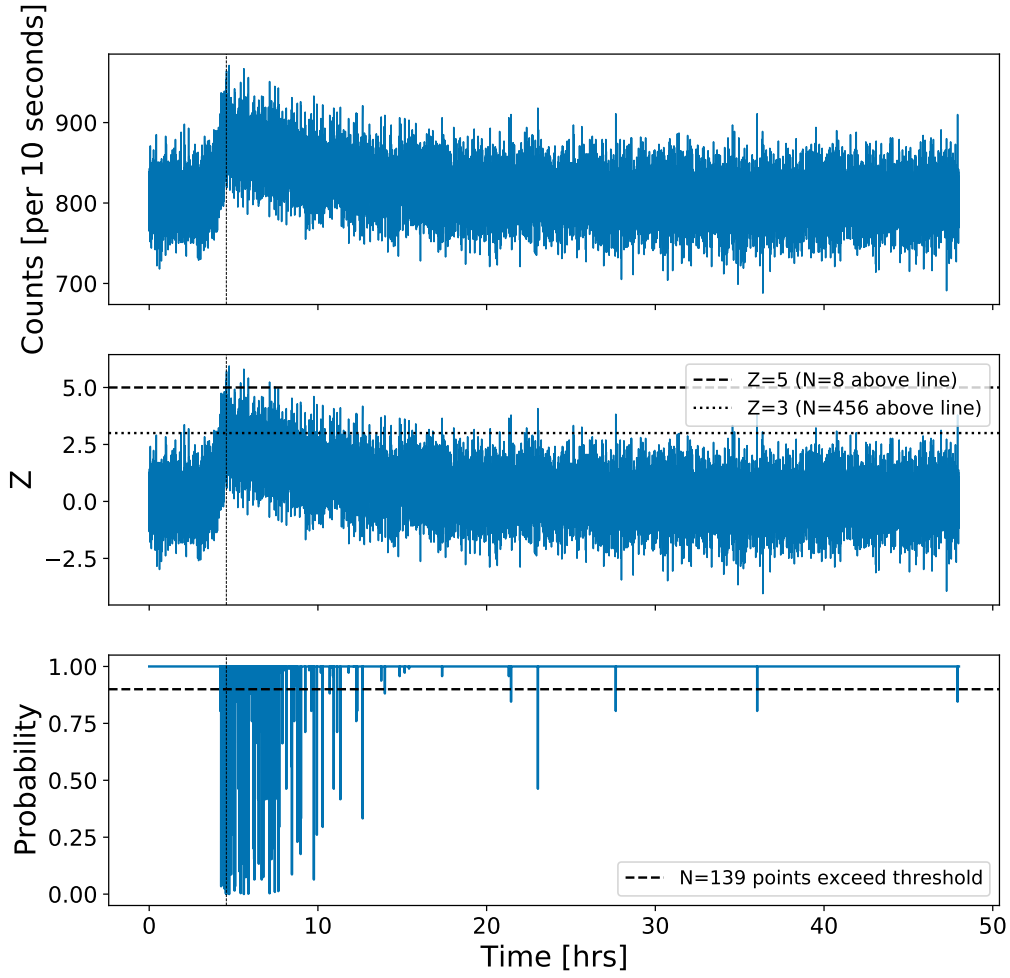


Figure 1.16: Single realisation of a simulation with a 10% GLE magnitude. Top panel shows the raw, artificial data from the simulation, with a 10-second cadence. Middle panel shows the number of measurements exceeding the  $Z = 3$  and  $Z = 5$  thresholds. The  $Z = 3$  and  $Z = 5$  thresholds are 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.

significant, and it is 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, as within two days of artificial 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 there being a true event if the number of significant measurements exceeds  $\sim 33$ ,  $\sim 6$ , and  $\sim 2$ , 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  $\sim 2 - 4$ . 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.

In the data, we can use the number of statistically significant measurements to warn whether there exists a possible event in the data, and closer inspection of the statistics tests can verify or exclude the existence of a true event. For example, in Figure 1.16, there are 139 points exceeding the 10% binomial limit, and 456 and 8 point above the  $Z = 3$  and  $Z = 5$  limits, respectively. This indicates the existence of an event in the data. On closer inspection it is clear from the grouping of the number of points exceeding each of the thresholds where the event occurred.

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 1.17 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 shows 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 1.17 that in the case where no averaging of data is 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

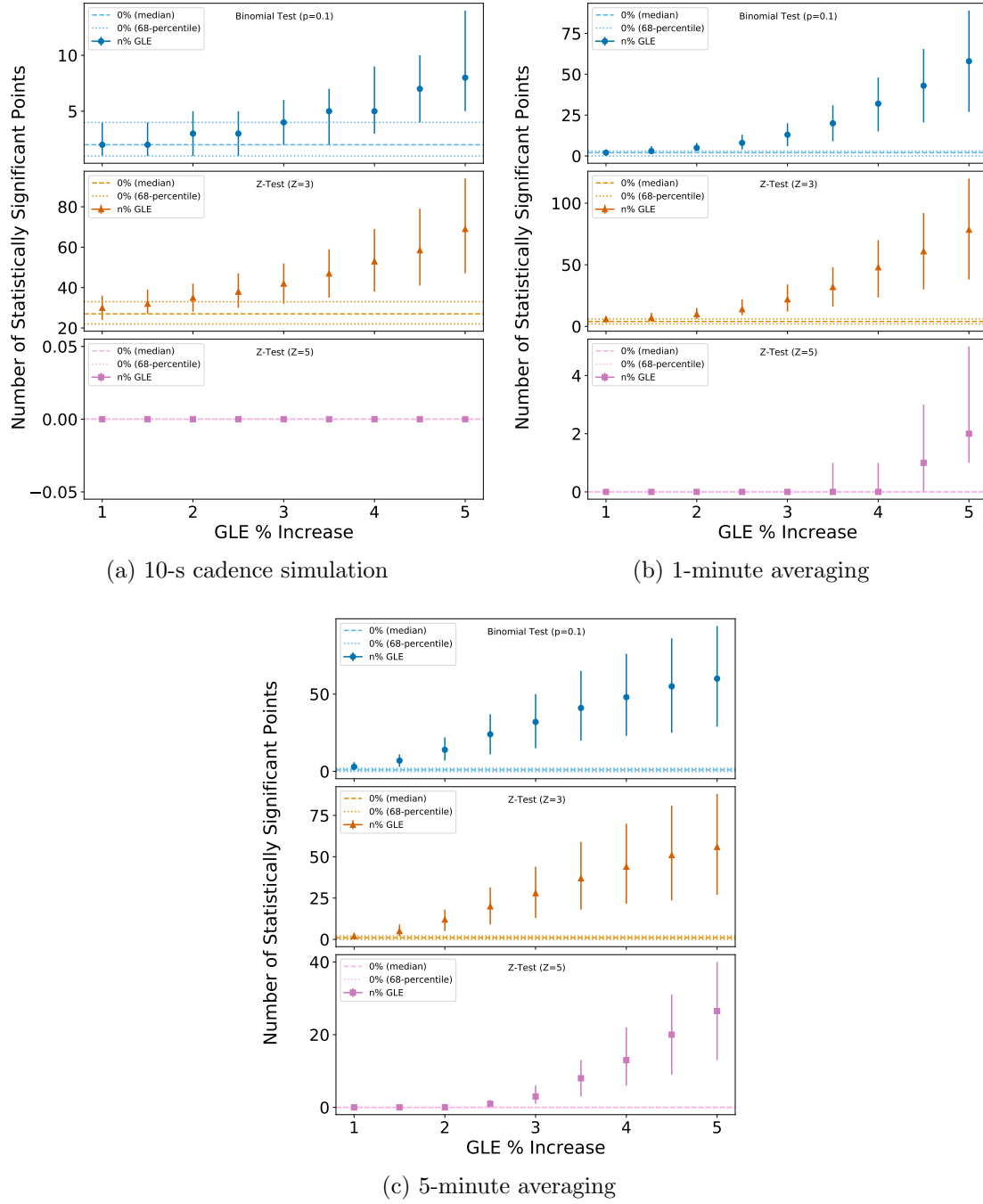


Figure 1.17: Summary plots showing the number of statistically significant measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) shows the results for the 10-s cadence data; (b) 1-minute averaged data; (c) 5-minute averaged data. 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. 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.

$> 4.5 - 5.0\%$ . To be able to see the results clearly, we only show the results for GLE magnitudes of up to  $5\%$ , and in Figure 1.17a, 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  $\geq 7.5\%$ .

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  $\sim 1.0 - 1.5\%$ , respectively.

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  $1 - 2\%$ . 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.

Overall, based on the values predicted in Section ??, 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).

### 1.6.4 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. 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 A for details) to determine the magnitude of GLEs that may be observed with a network of detectors using this configuration. On overarching assumption in this multi-station analysis is that the detectors are all geographically close, such that the delay in the induced signals are 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  $\sim 15$  km from University of Birmingham; however, with the exception of the existing University of Birmingham station 14001, which is a 4-detector station, each of the stations in the Birmingham node is a 2-detector station. This means that there are in total an additional 14 individual detectors in the Birmingham node.

Table 1.4: ... with no injected GLEs ...

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

[discussion and plots on increasing the number of stations for different GLE magnitudes... need a summary plot to show this, i.e. stats test on the average

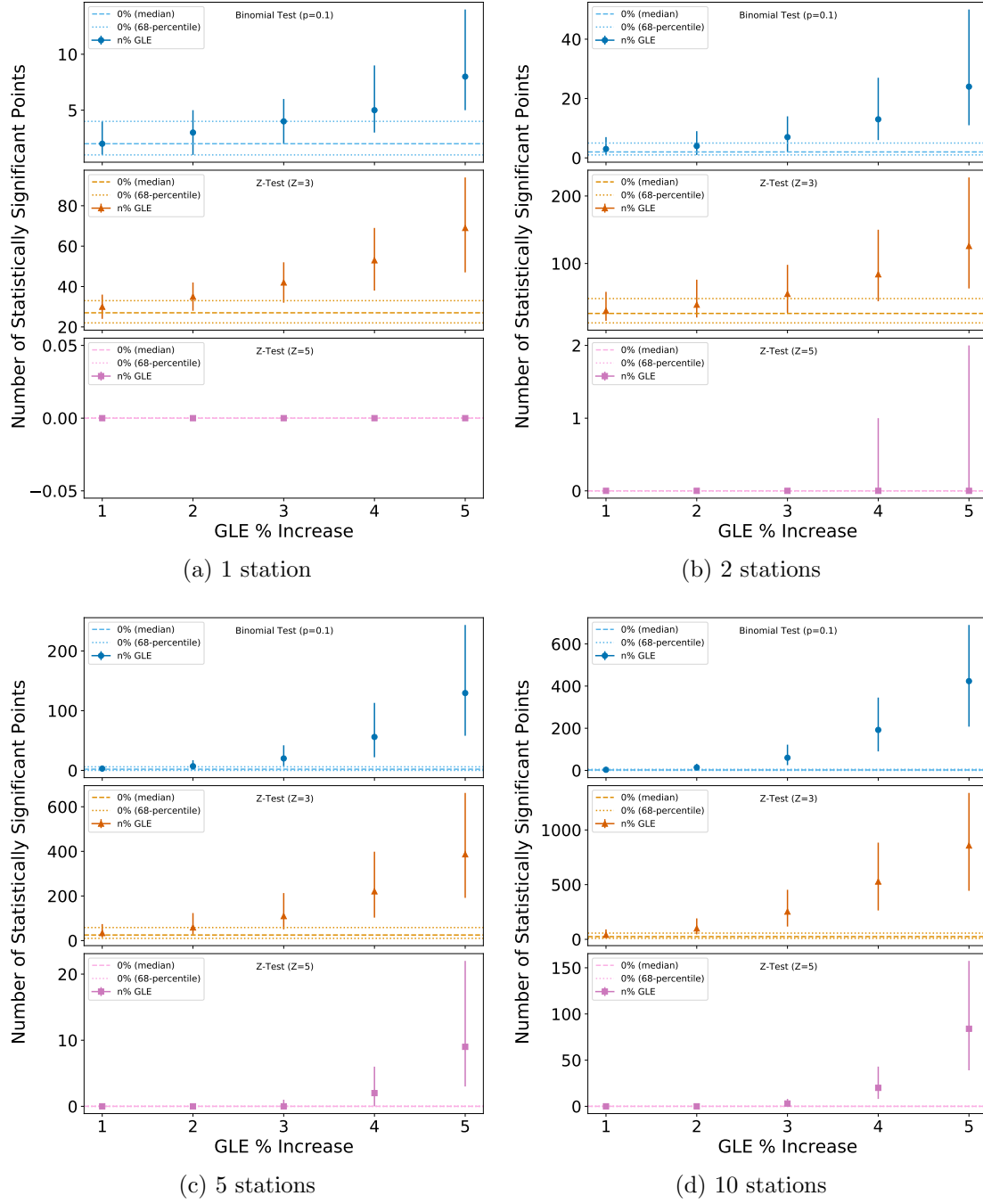


Figure 1.18: Summary plots showing the number of statistically significant measurements using the binomial- and Z-tests on the simulations of artificial data with varying magnitudes of GLEs injected. (a) ...; (b) ...; (c) ...; (d) ... . In each plot, the top panel shows the number of significant points 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. 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.

timeseries for 2, 5, 10 stations (also show single stations from above!!)...

[discussion on the spread being a side-effect of the randomly sampled rise and decay, allowing for more measurements to remain above the threshold for longer if a longer decay...]

In addition, we also perform cross-correlation analyses between the multiple stations simulated. This again relies 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, to also investigate the effect that this has on our ability to detect the event using this method.

[Section on CCF and need to show how the effects of varying the decay time varies the ability to observe - also show how ability to claim varies with number of stations used... this is for 2 per cent GLE...]

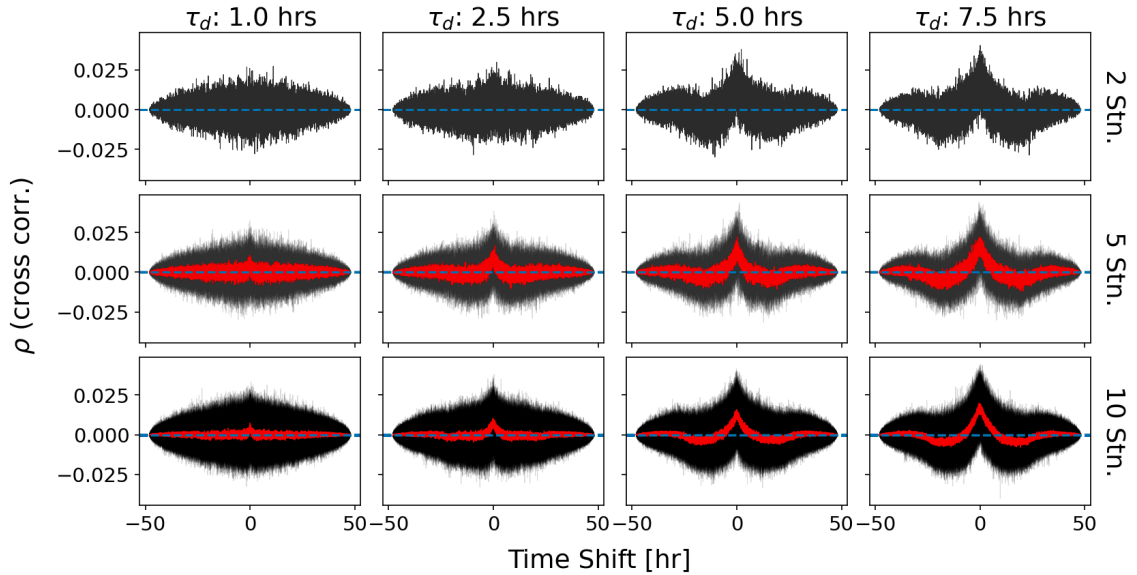


Figure 1.19: ... for 2% GLE magnitude ...

We see from Figure... that increasing the number of stations means we can average over the CCFs which results in a less-noisy CCF. This is advantageous as it allows us to more clearly detect the correlated signals between 5 and 10 stations, versus with only 2 stations... There is not a significant benefit at this level of GLE

in increasing from 5 to 10 stations. However, the benefit of an increased number of stations is that we are able to reduce the noise on the combined signal, which allowed us to observe significant peaks in the statistics tests (in earlier fig...)...

[Also show CCF for 1 per cent GLE with varying the number of stations (do we also want to show the varying of the decay time too? basically a repeat of the above plots but for 1 per cent GLE???)]

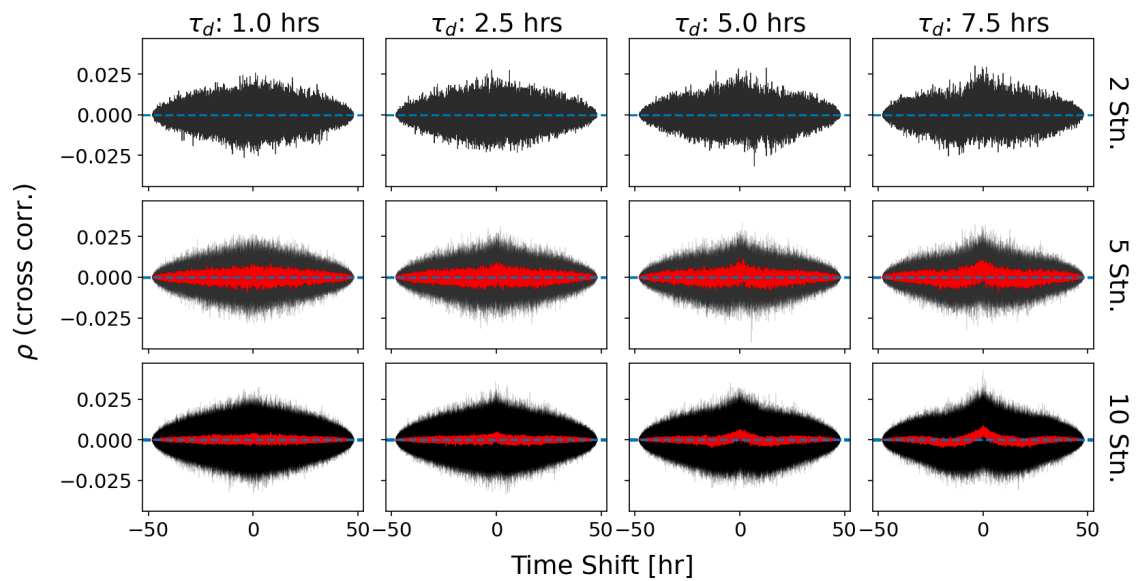


Figure 1.20: ... for 1% GLE magnitude ...

[Do we mention PCA??? maybe no...]

[conclusive statements... how many stations do we think should be changed?]

## 1.7 Conclusions

We have presented a new HiSPARC station configuration and investigated its performance as a monitor of space weather events...

...

...

We leave the reader with the following points:

- 1.

- 2.
- 3.
4. ... It is therefore important that this configuration of HiSPARC detector is maintained until at least 2026, to ensure a complete study is performed.



# A Simulations of the Ground Level Enhancements for the Analysis of HiSPARC Station 14008

## A.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
- $\lambda$ : Background count level
- $\sigma$ : Detector noise level
- $A$ : Percentage amplitude of GLE increase
- $\tau_r$ : Rise time of GLE

- $\tau_d$ : Decay time of GLE

Using these variables, the mathematical form of the model was expressed by:

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

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 A.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.

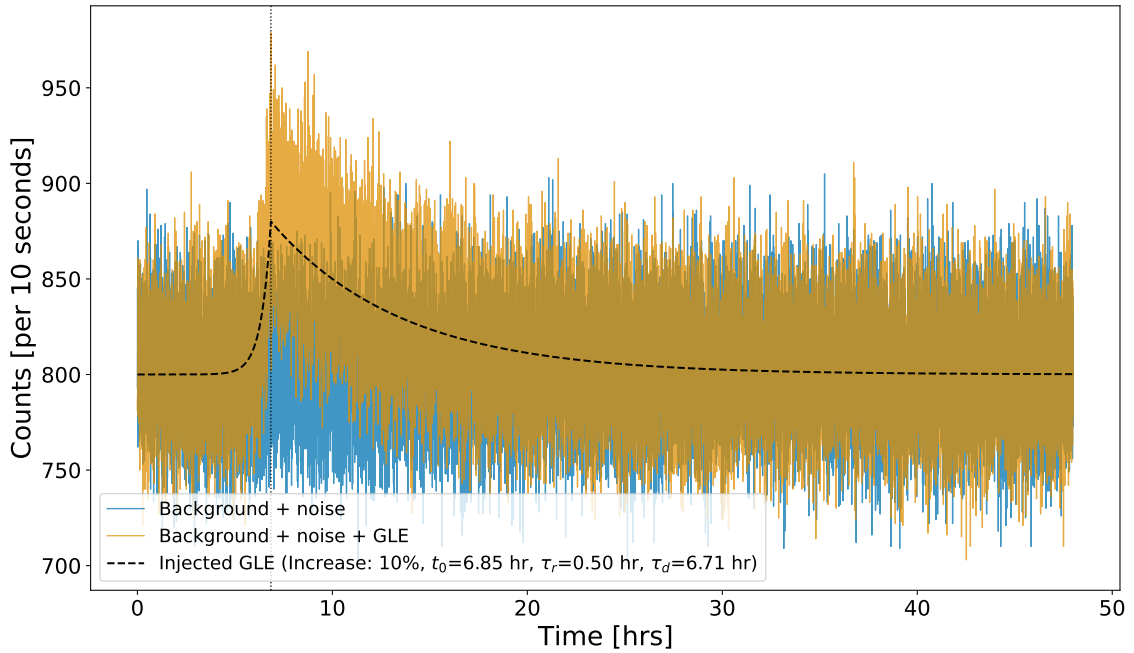


Figure A.1: ...

## A.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 A.2.

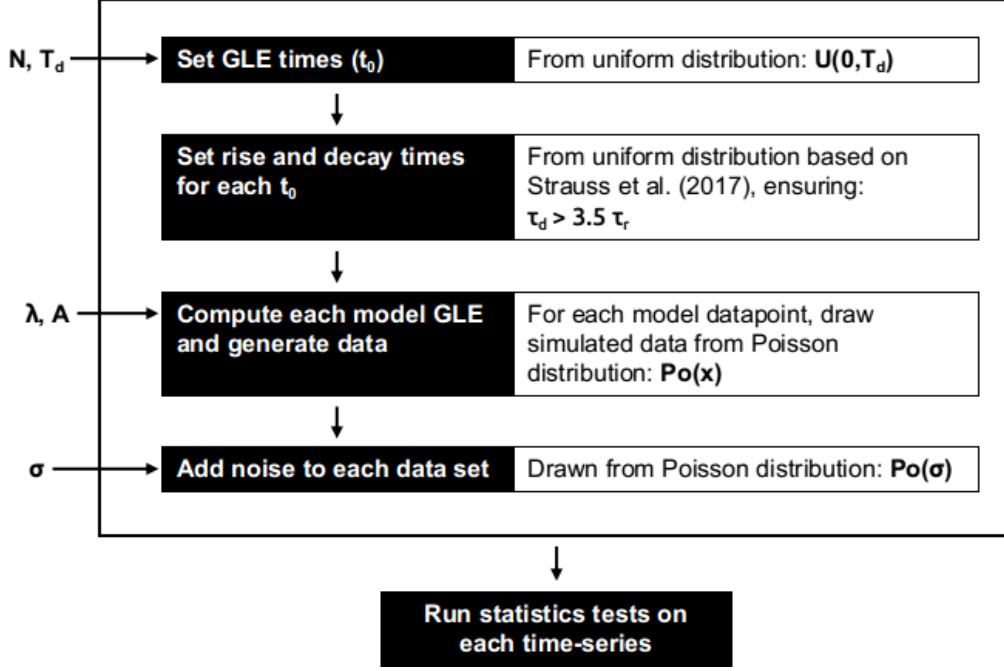


Figure A.2: Flowchart showing the step-by-step processes in the generation of the artificial HiSPARC 14008 GLE time series data.

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,  $\mathcal{U}(0, 75)$  and  $\mathcal{U}(15, 500)$ , based on the information from [Strauss et al. \(2017\)](#), whilst also ensuring  $\tau_d > 3.5 \tau_r$ , ([Strauss et al., 2017](#)).

With these properties draw, and the user inputting the background level ( $\lambda$ ), noise-level ( $\sigma$ ), and percentage increase of the GLE ( $A$ ), each individual GLE was simulated according to equation (A.1).

After the creation of the simulated data, statistical tests were performed to determine whether the GLE is observed.

[!!! additional bit for multiple stations simulations... basically repeat the same, but do select the  $A$  and  $\lambda$  from small-width normal distribution around values provided by user - to simulate ever so slight differences between the different stations set-ups...]

# Bibliography

- Blackmore E. W., Stukel M., Trinczek M., Slayman C., Wen S., Wong R., 2015, [IEEE Transactions on Nuclear Science](#), 62, 2792
- Cecchini S., Sioli M., 2000, arXiv:hep-ex/0002052
- Dubey A., Kumar S., Dubey S. K., 2016, [Journal of Physics: Conference Series](#), 755, 012056
- 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>
- McIntosh S. W., Chapman S., Leamon R. J., Egeland R., Watkins N. W., 2020, [Solar Physics](#), 295, 163
- 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/](https://www.nikhef.nl/pub/services/biblio/theses_pdf/)
- NMDB 2018, NMDB Event Search Tool (NEST), <http://www.nmdb.eu/nest/>
- Particle Data Group et al., 2020, [Progress of Theoretical and Experimental Physics](#), 2020
- 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
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, [PeerJ Comput. Sci.](#), 2, e55
- Strauss R. D., Ogunjobi O., Moraal H., McCracken K. G., Caballero-Lopez R. A., 2017, [Solar Physics](#), 292, 51
- Thomas S., Owens M., Lockwood M., Owen C., 2017, [Ann. Geophys.](#), 35, 825
- 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