

BERGISCHE UNIVERSITÄT WUPPERTAL
FAKULTÄT 4
MATHEMATIK UND NATURWISSENSCHAFTEN
FACHGRUPPE PHYSIK

Correlation of Solar Wind velocity and Scaler data of the Pierre Auger Observatory

Master Thesis
Daniel Rosenbaum

September 26th, 2017

Markers:

Prof. Dr. rer. nat. Karl-Heinz Kampert
Prof. Dr. rer. nat. Christian Zeitnitz

Abstract

In this Master thesis the anticorrelation between Solar Wind velocity and the variation in the count rate of the Pierre Auger Observatory's low energy cosmic ray 'Scaler' mode is studied. This anticorrelation was predicted theoretically by the diffusion-convection mechanism and has already been measured by the GRAPES-3 experiment. To measure this anticorrelation, several unrelated effects like the solar rotation and the solar magnetic cycle are corrected and periods of extreme solar activity, as during Forbush decreases, are removed from the analysis. Additionally, several detector-specific corrections are applied to the Scaler data. The anticorrelation is then fit with a linear model and yields a slope of $[-6.8 \pm 0.3 \text{ (stat.)} \pm 0.7 \text{ (syst.)}] \cdot 10^{-4} \% \text{ s km}^{-1}$ for the fractional change of the Scaler rate, with respect to the Solar Wind velocity. The linear model describes the data well and a constant model can be rejected with a significance larger than 3 standard deviations.

Contents

1. Introduction	1
2. Cosmic rays and their detection at the Pierre Auger Observatory	3
2.1. Cosmic rays	3
2.2. Extensive air showers	5
2.3. Auger Hybrid Detector	5
3. Solar Wind	8
3.1. Periodic effects	8
3.2. Forbush decreases	9
3.3. Theoretically expected anticorrelation of Solar Wind velocity and cosmic ray rate	10
4. Data description	12
4.1. Scaler data	12
4.1.1. Corrections to Scaler data	13
4.1.2. Scaler dataset A	16
4.1.3. Scaler dataset B	16
4.2. NASA OMNIWeb data	18
4.2.1. Input data preparations and cleaning	18
4.2.2. Time shifting	19
4.2.3. Building 1-minute averages	19
5. Simple analysis with few corrections	20
5.1. GRAPES-3 experiment	20
5.2. GRAPES-3 data treatment	20
5.3. Comparison of GRAPES-3 muon intensity and Scaler data results	22
6. Independent analysis of the anticorrelation	27
6.1. Determining a time-window	27
6.2. Scaler data treatment	34
6.2.1. Removing transient effects	34
6.2.2. Removing daily patterns	36
6.3. Minimal required entries	38

Contents

6.4. Variance analysis	41
6.5. Results	43
7. Conclusion and outlook	52
A. Appendix	53
A.1. Determining a time window	54
A.2. Scaler data treatment	62
A.3. Variance analysis	64
A.4. Results	65
References	75

1. Introduction

Solar magnetic activity influences the transport of charged particles in the interplanetary medium within the heliosphere. Thus, the rate of cosmic rays penetrating the Earth's atmosphere is affected by the Sun. In the past these effects have been studied using neutron monitors, but today, muon detectors and the low-energy modes of high-energy observatories also can be used to study the impact of space weather on cosmic rays. The diffusion-convection mechanism in the heliosphere predicts an anticorrelation of cosmic ray intensity with changes in Solar Wind velocity. Recently, the *GRAPES-3* experiment has published an observation of the anticorrelation of muon intensity and Solar Wind velocity [1]. Through this work, this anticorrelation has now been validated with data from the low-energy 'Scaler' mode of the surface detector of the Pierre Auger Observatory.

In the following chapter, the basics about cosmic rays will be covered and important past observations will be briefly stated. What an extensive air shower is and how the Hybrid Detector of the Pierre Auger Observatory detects the secondary particles that are created when cosmic rays interact with the atmosphere will also be explained.

In Chapter 3, the most important characteristics of the Solar Wind and the theoretical motivation for the anticorrelation of Solar Wind velocity and cosmic ray rates will be briefly covered.

In Chapter 4, the datasets used in this analysis, the Solar Wind velocity data provided through NASA OMNIWeb, and two different Scaler datasets from the Pierre Auger Observatory, will be described. The corrections applied needed to provide the data quality required for this study are also included in this chapter.

A duplication of the analysis used for the *GRAPES-3* result is then performed on a preprocessed version of the Scaler dataset in Chapter 5. The analysis of this yields results comparable to the *GRAPES-3* study.

Finally, in Chapter 6, an independent analysis is performed on a new Scaler dataset that was also corrected for long-term effects that would otherwise become visible as data from a 10-year timespan is used. First, a fitting averaging time-window for both the Scaler and Solar Wind velocity data is determined. It filters the noise on small time scales while preserving the larger scale features. Afterwards, additional corrections are applied to the

1. Introduction

Scaler data to remove daily oscillations, seasonal patterns, weather effects and detector instabilities. The anticorrelation of Scaler data and Solar Wind velocity data is then fit and the results from both the preprocessed Scaler data and the GRAPES-3 study are verified.

2. Cosmic rays and their detection at the Pierre Auger Observatory

This chapter features a brief explanation of what cosmic rays are and how astroparticle physics developed in history. In 1912 Victor Hess was able to show that the flux of ionizing radiation increases with an increasing altitude above 300 m. This led to the conclusion that this radiation must originate from space. This was the beginning of the study of cosmic rays. 15 years later Dimitry Skobelzyn made cosmic rays visible in a cloud chamber. It was in 1938 when Werner Kolhörster and co-workers [2] and Pierre Auger and co-workers [3] discovered that high-energy cosmic rays produce extensive air showers in the Earth's atmosphere. This made their analysis possible for the first time in history and laid the foundation for many modern astroparticle experiments. In the following sections it will be explained how cosmic rays interact with the Earth's atmosphere and how the Pierre Auger Observatory in Argentina was designed to measure cosmic rays.

2.1. Cosmic rays

Cosmic rays are high-energy radiation from space which hit the Earth. They consist mostly of ionized nuclei – most prominently protons, but also nuclei of elements with greater mass – and electrons. These particles can have energies up to more than 10^{20} eV making their study extremely interesting. For comparison the current world record for particle accelerators is held by the LHC with $6.5 \cdot 10^{12}$ eV [4]. Many Beyond Standard Model theories only produce predictable results for very high energies, which cosmic rays might provide.

Mid-range energy cosmic ray particles stem from within the Milky Way and may originate from supernova explosions. After being thrown out by such a dying star, the particles are further accelerated by several mechanisms, e.g. the first order Fermi statistic acceleration from astrophysical shock waves or the second order Fermi acceleration from randomly moving interstellar magnetized clouds [5]. Still, so-called ultra-high-energy cosmic rays (UHECR), cosmic rays with energies above 10^{18} eV, are so far inexplicable with no known acceleration sites capable of reaching these energies.

Although the possible high energies of cosmic rays are promising at first sight, their study is hindered by the dependence of the particle flux on the energy: Generally, the flux decreases with increasing particle energy according to a power law as

$$\frac{dN}{dE} \propto E^{-\gamma}, \quad (2.1)$$

with varying spectral index γ . The energy dependence of the flux of cosmic rays is depicted in Figure 2.1.

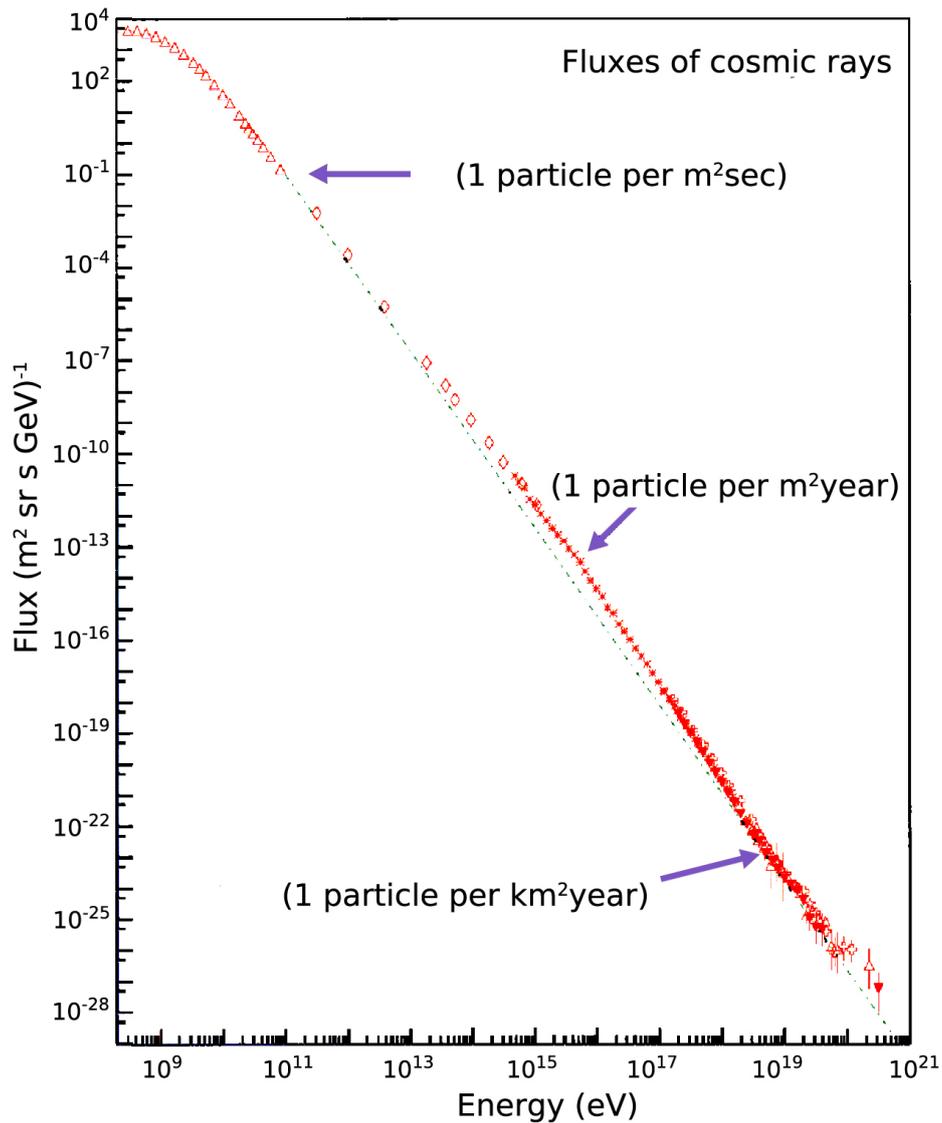


Figure 2.1: The flux of cosmic rays as a function of the energy [6].

2.2. Extensive air showers

Cosmic rays with low energies can be measured directly with detectors on high-flying balloons or on satellites. As the flux-diagram in Figure 2.1 shows, this gets more problematic for higher energies of incident particles. One would either have to wait for a long time or build huge structures with large areas in space.

When a high-energy cosmic ray – the so-called *primary particle* – enters the Earth’s atmosphere it will almost certainly interact with air nuclei and initiate a cascade of particle interactions and secondary particles, a so-called *extensive air shower*, illustrated in Figure 2.2. This shower consists of different components:

- electromagnetic component: neutral pions decay into photons and produce pairs of electrons and positrons. These then emit photons via bremsstrahlung. In addition to that, muons produced in the shower can decay into electrons which then follow the above process.
- hadronic component: hadronic interactions result in protons, neutrons or other hadrons like pions and kaons.
- muonic component: pions and kaons decay into muons.
- neutrino component: neutrinos from decays of charged pions or leptons.

Secondary particles hitting the ground can be measured with arrays of particle detector stations. Additionally, the UV fluorescence light produced when charged shower particles interact with atmospheric molecules can also be detected with telescopes.

2.3. Auger Hybrid Detector

The Pierre Auger Observatory experiment is named after the discoverer of extensive air showers Pierre Victor Auger (1899-1993). It is located in the Pampas near Malargüe, Argentina (35°S, 69°W, 1400 m a.s.l.) and was designed to study cosmic rays at the highest energies. It is a *Hybrid Detector* as it provides two independent methods to measure high-energy cosmic rays for cross-calibration. The structure of the Pierre Auger Observatory is shown in Figure 2.3.

Of primary importance to this work is the *Surface Detector Array*, which consists of 1660 water-Cherenkov detector stations, shown in Figure 2.4, arranged in a 1500 m spaced triangular grid covering a total area of 3000 km². The stations are filled with purified water which is monitored by three photomultipliers (PMTs) in order to detect the Cherenkov-radiation that is produced by the shower particles moving at a speed exceeding the speed of light in water. The amount of light in all stations allows a determination

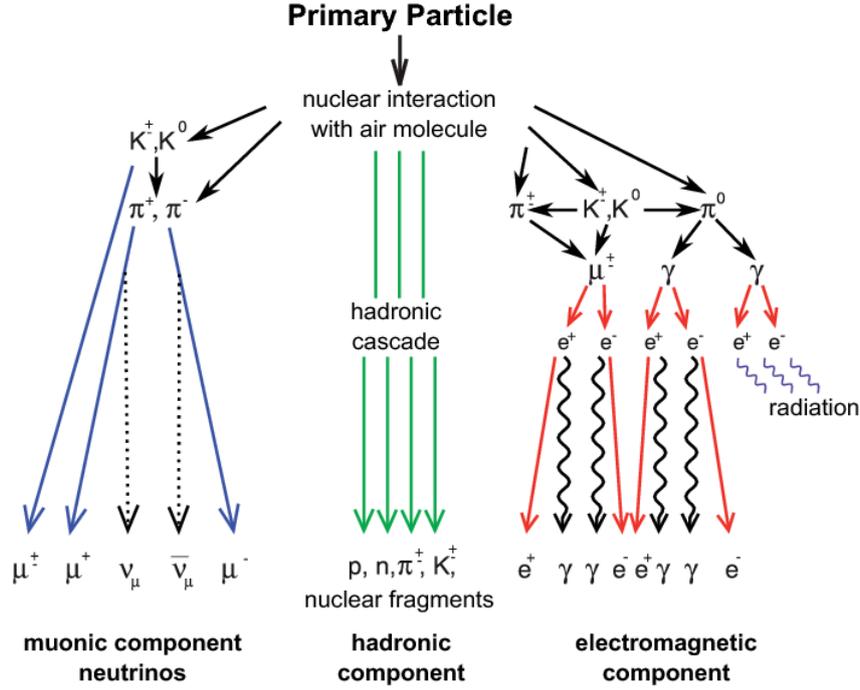


Figure 2.2: Development of an extensive air shower, caused when a high-energy primary particle interacts with the atmosphere [7].

of the energy of the primary cosmic ray particle and the detection times at different detector stations make a reconstruction of the cosmic rays' trajectory possible. In this analysis only data from the Surface Detector Array will be used.

Secondly, there is the *Fluorescence Detector* composed of 27 fluorescence telescopes housed at four different locations at the edges of the Surface Detector array. In moonless nights these optical detectors can detect fluorescence light produced by collisions of charged particles of an air shower with atmospheric nitrogen. This observation allows for the reconstruction of the direction of the cosmic rays and a nearly direct measurement of the total energy of the primary particle via the total amount of light emitted from the air shower.

2. Cosmic rays and their detection at the Pierre Auger Observatory

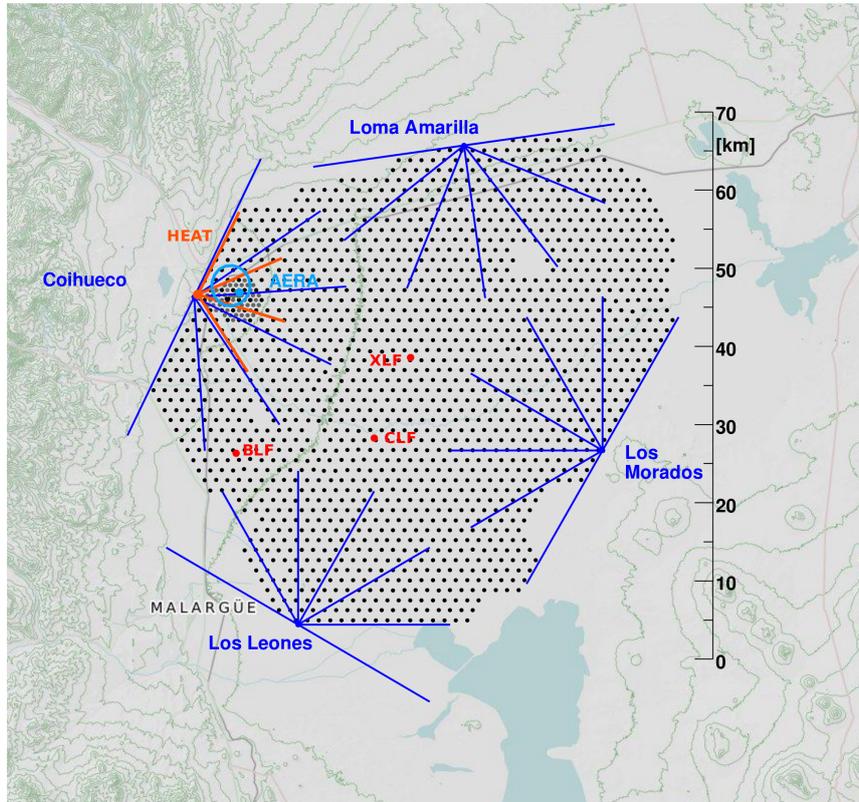


Figure 2.3: Structure of the Pierre Auger Observatory: fluorescence detectors (blue lines) and surface detector stations (black dots) [8].



Figure 2.4: A surface detector station of the Pierre Auger Observatory [9].

3. Solar Wind

Solar Wind is a supersonic stream of charged particles emitted from the Sun. Mainly it is comprised of protons, electrons and alpha particles with energies in the keV range and velocities of several hundred kilometres per second. The wind originates from the thermal expansion of the solar corona against the gravitational pull of the Sun. Via this ejection mechanism the Sun loses about $3 \cdot 10^{-14} m_{\odot}$ per year [10], which is equal to $\sim 10^9$ kg/s. In figure 3.1 one can see how the Solar Wind interacts with the Earth's magnetosphere.

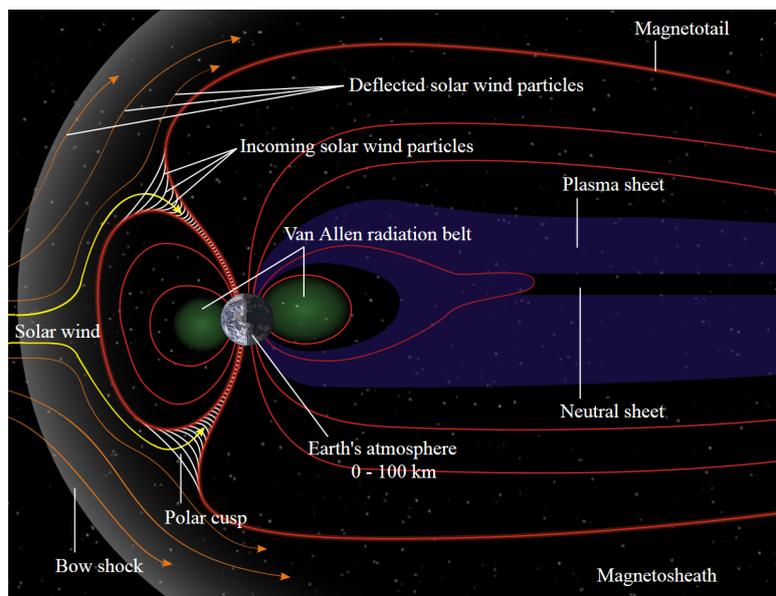


Figure 3.1: Schematic view of the Solar Wind hitting the Earth's magnetosphere [11].

3.1. Periodic effects

The Sun does not eject the Solar Wind uniformly across its surface, therefore *solar rotation* influences Solar Wind intensities at Earth. As the Sun consists of gaseous plasma, its rotation speed varies with latitude and is fastest at the equator. To unify the

studies of solar rotation *Bartels' Rotation Number* was defined. It is a serial count of the apparent rotations of the Sun since November 9th 1853 as viewed from Earth with a defined length of exactly 27 days, which is close to the synodic Carrington rotation rate of 27.2753 days.

In addition to that, there is also the so-called *solar magnetic activity cycle*, which is the roughly 11-year periodic change of the Sun's activity, mainly characterized by the number and size of sunspots. Each such cycle features a solar minimum and maximum, which can be measured in long-term studies of neutron monitors and particle detectors and is illustrated in Figure 3.2.

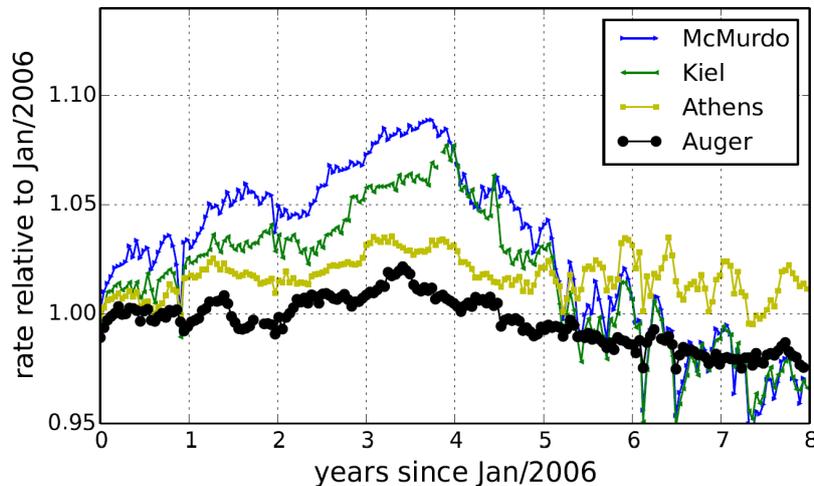


Figure 3.2: 15-day averages of AoP - and pressure-corrected Scaler data (the necessary quantities will be explained later in Section 4.1) compared with different neutron monitor data from McMurdo, Kiel and Athens showing the global peak at the minimum solar period [12].

3.2. Forbush decreases

A *Forbush Decrease* is a transient event characterized by a strong short-term depression of the cosmic ray flux observed on Earth. Forbush decreases can be observed and identified by neutron monitors and muon detectors; their structure is asymmetrical with a sudden start and a smooth recovery which typically takes several days [13]. These decreases are caused by coronal mass ejections of the Sun: large bursts of matter and magnetic disturbances, that sweep away cosmic radiation in the heliosphere. In Figure 3.3 an example for a Forbush decrease measurement of the Los Cerrillos neutron monitor in correlation with the Scaler rate of the Pierre Auger Observatory (that will be explained in Section 4.1) is presented.

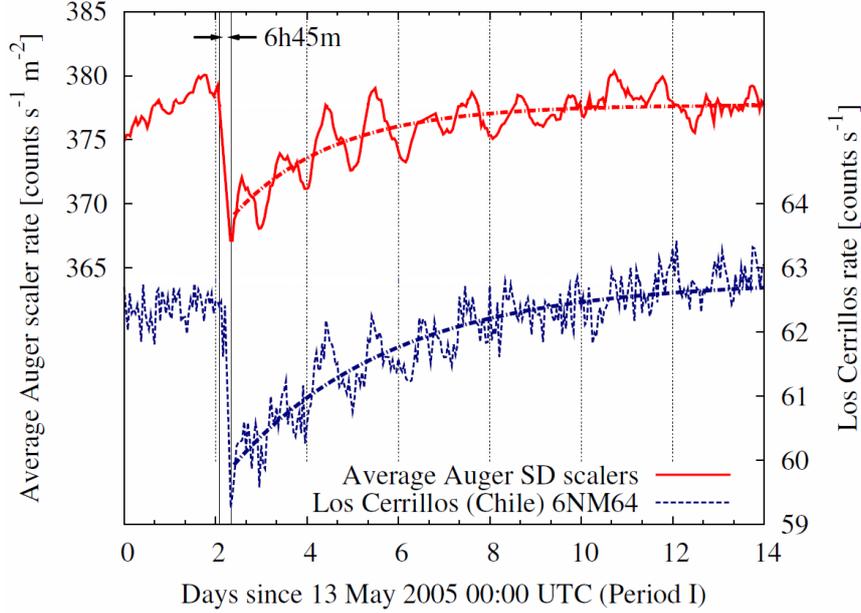


Figure 3.3: Comparison of the pressure-corrected Scaler rate (red solid line, for a detailed explanation of the Scaler rate and its pressure correction see Section 4.1) with the Los Cerrillos neutron monitor rate (blue dashed line). It takes about 6 hours and 45 minutes to reach the minimum and about 9 days to fully recover from the decrease. The recovery periods of both observables have been fitted exponentially (red and blue dot-dashed lines), which describes the data well [13].

3.3. Theoretically expected anticorrelation of Solar Wind velocity and cosmic ray rate

The flux of energetic charged particles in interplanetary space can be described by the diffusion-convection equation [14]:

$$\frac{\partial f}{\partial t} = \underbrace{\vec{\nabla} \cdot (D \vec{\nabla} f)}_{\text{diffusion}} - \underbrace{\vec{\nabla} \cdot (\vec{v}_{\text{SW}} \cdot f)}_{\text{convection}} \quad (3.1)$$

where

- f denotes the variable of interest, e.g. the concentration of cosmic rays,
- D is the diffusion coefficient, which is temperature dependent, and
- \vec{v}_{SW} is the Solar Wind velocity.

It can be derived from the continuity equation:

$$\frac{\partial f}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0 \quad (3.2)$$

where the total flux \vec{j} has two sources:

1. the diffusive flux $\vec{j}_{\text{diffusion}}$, which can be approximated by Fick's first law [15] as $\vec{j}_{\text{diffusion}} = -D\vec{\nabla}f$ stating the flux of diffusing material is antiproportional to the local concentration gradient and
2. the advective flux $\vec{j}_{\text{advection}} = \vec{v} \cdot f$ which represents an overall convection of the material.

Substituting the total flux $\vec{j}_{\text{total}} = \vec{j}_{\text{diffusion}} + \vec{j}_{\text{advection}}$ into the continuity Equation (3.2) directly yields above convection-diffusion Equation (3.1). The latter equation directly shows that the concentration, and by that also the rate, of cosmic rays should be antiproportional to changes in the Solar Wind velocity.

4. Data description

As mentioned in chapter 1 the low-energy modes of astroparticle detectors can be used to study the behaviour of galactic cosmic rays, especially their solar modulation. While GRAPES-3 [1] uses a large-area muon detector, this analysis will use the low-energy mode of the Surface Detector of the Pierre Auger Observatory, the so-called ‘Scaler’ mode. For the Solar Wind data, both analyses use the same source: the NASA OMNIWeb service, which uses data provided by three different satellites (see below).

4.1. Scaler data

On March 30th 2005 the *Scaler* mode was introduced for the Surface Detector array of the Pierre Auger Observatory. Initially, it was defined as any signal above 3 counts of the analog-to-digital converter (ADC) above the baseline of at least one of the three PMTs. This corresponds to an energy threshold of about 15 MeV and yields an average rate over the complete array of about 3.6 kHz per detector. This rate is read out every second and sent to the Central Data Acquisition System to be stored in compressed files [16]. As implemented there were two limitations:

1. At the beginning of each second, the rate is read and reset responding to a trigger with about 100 Hz, which leads to a counting fluctuation of about 1 %, and
2. every 61 seconds, the local station is calibrated, which has a higher priority than the Scaler reading process, resulting in pairs consisting of a ‘longer’ second (with a spike in the rate) and a ‘shorter’ second (with a lower than average rate) around the time of the calibration process.

To address these limitations, on September 20th 2005 the timing issues were corrected via a software update and the Scaler rate was also redefined to count rates above 3 ADC counts but below 20 ADC counts. This corresponds to an energy limit of about 100 MeV. This was done to diminish the sensitivity to muon signals, which act as background noise for the search for Gamma-Ray-Bursts (GRBs). These GRBs are of interest for several studies, which is why the Scaler rate was redefined. From these changes, the average rate over the array drops to about 2 kHz per detector [17]. In this analysis, only data from the latter period with the upper threshold is used.

For later use, some helpful quantities regarding the Scaler rate are defined below:

- S_{id}^{phys} : The Scaler rate of the surface detector station with identification number id
- $\mathcal{S}^{\text{phys}} = \langle \langle S_{id}^{\text{phys}}(t) \rangle_{\forall id} \rangle_{\forall t}$: The whole (considered) array and time averaged Scaler rate
- $S_{id} = S_{id}^{\text{phys}} / \mathcal{S}^{\text{phys}}$: The Scaler rate of the surface detector station with identification number id normalized by $\mathcal{S}^{\text{phys}}$

4.1.1. Corrections to Scaler data

Scaler data is affected by random noise. This can be dealt with using RMS cuts, by setting a minimum number of detector stations contributing to the average rate and by accounting for systematic offsets that are explained in the following.

Pressure correction

It has been shown that the Scaler rate is correlated with atmospheric pressure, with a change of $-3.6 \pm 0.2(\text{stat}) \%$ per hPa [13]. An example highlighting this correlation is shown in Figure 4.1.

Area-over-peak correction

The area-over-peak (AoP) value is defined as the ratio of the deposited charge VEM_q to the peak VEM_p of pulses resulting from the passage of a single vertical muon equivalent in the water-Cherenkov detector, i.e.

$$AoP = \frac{VEM_q}{VEM_p} \quad . \quad (4.1)$$

The typical signal of a single vertical muon in a water-Cherenkov detector starts with a rapid increase (the *peak*) followed by an exponential decay. The decay is the result of, one or many, reflections of the Cherenkov light by the inner walls of the detector station. Obviously, the length of the decay (longer decay $\hat{=}$ larger *area*) depends on the transparency of the water, as well as the reflectivity of the inner walls of the detector.

As expected there is a linear correlation between AoP and the Scaler rates of individual detector stations, which is depicted in Figure 4.2. A larger AoP and therefore larger „tail“ in a water-Cherenkov detector signal produces more counts above the baseline of the PMT's ADC. A linear function like Equation (4.2) was fit to the data and by using

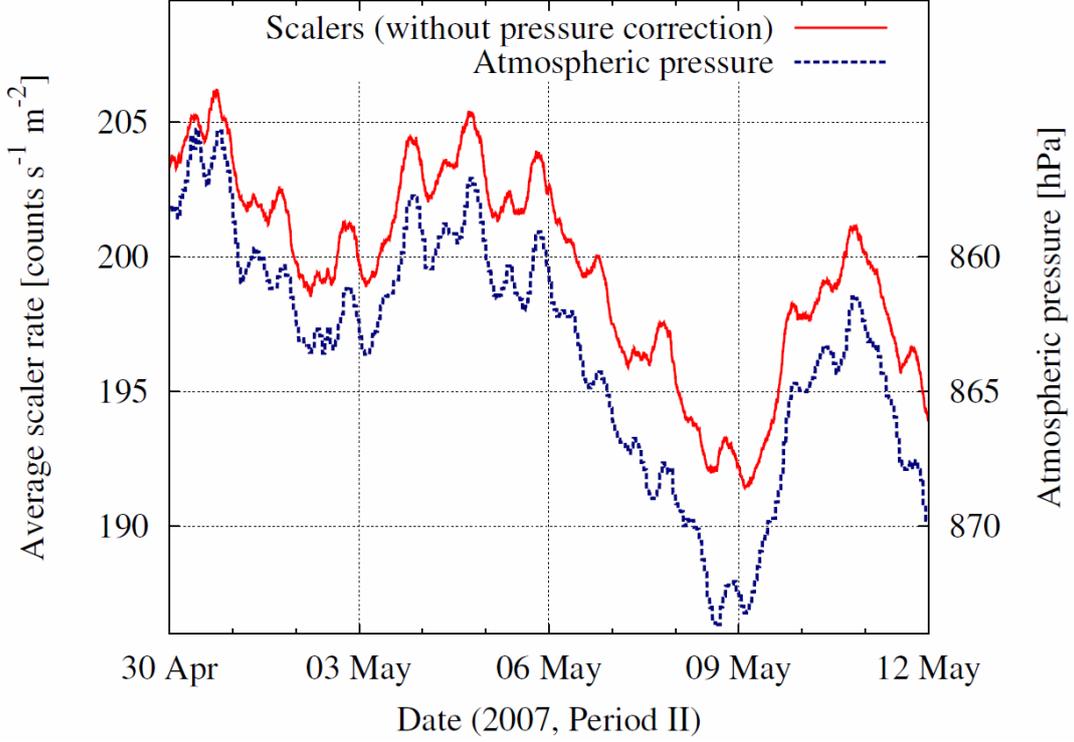


Figure 4.1: Scaler rate averaged over time bins of 5 minutes and over all detectors in the array (solid red line); and atmospheric pressure (dashed blue line and reversed scale) for the first 12 days of May 2007 [13].

the resulting fit parameters a corrected Scaler rate for each station, $S_{id}^{\text{corr}/AoP}$, can be obtained according to Equation (4.3):

$$S_{id} = m^{S,AoP} \cdot AoP_{id} + b^{S,AoP} \quad (4.2)$$

$$S_{id}^{\text{corr}/AoP}(t) = S_{id}(t) - m^{S,AoP} \cdot (AoP_{id}(t) - \overline{AoP}) \quad (4.3)$$

where AoP_{id} is the AoP value for a water-Cherenkov detector station with identification number id and $\overline{AoP} = \langle \langle AoP_{id}(t) \rangle_{vid} \rangle_{vt}$ is the time average of AoP for the whole (considered) array.

Scaler rates can then be calculated by simply multiplying the corrected rate with the global average $\mathcal{S}^{\text{phys}}$. For the eight years of data between 2006 and 2013 according to [12] the slope of the linear fit was found to be:

$$m^{S,AoP} = 0.21 \pm 0.01 \text{ (stat.)} \quad (4.4)$$

4. Data description

The area-over-peak effect comes into play when looking at large time scales and is therefore important for this analysis. This effect is seen in the 8-year comparison of the uncorrected Scaler rate and the AoP -corrected Scaler rate in Figure 4.3. The non-constant behaviour, namely the peak in mid-2009, is caused by solar activity, which is the main interest of this study.

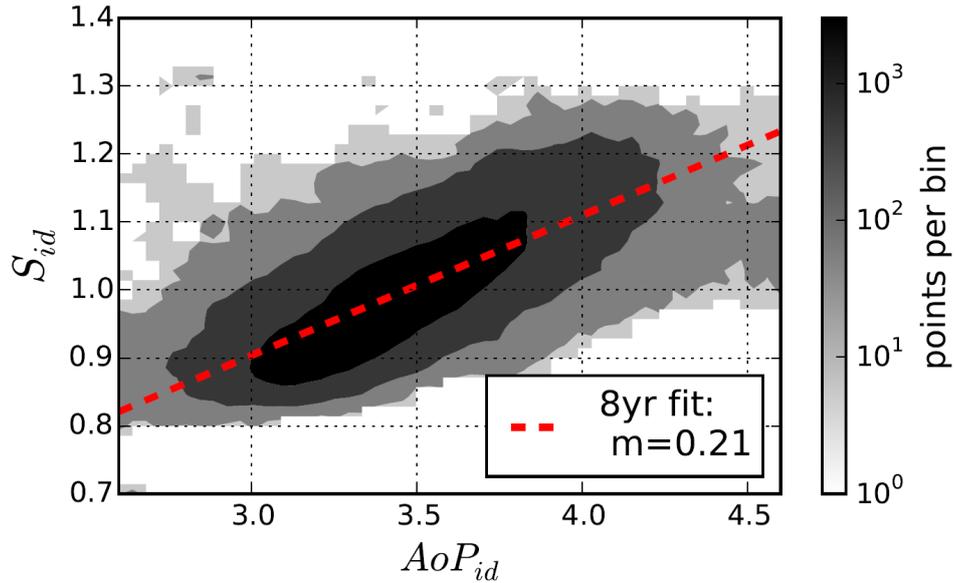


Figure 4.2: Two-dimensional histogram showing the correlation between the 5-days averaged Scaler rates S_{id} and AoP_{id} for data from individual WCD stations [12].

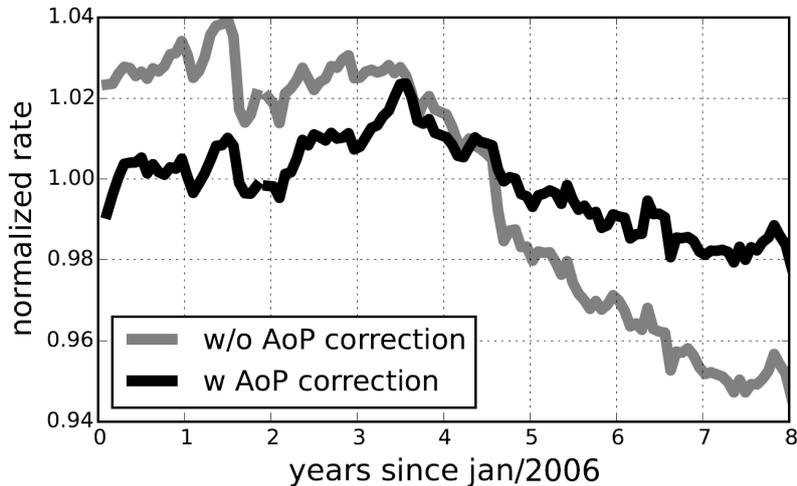


Figure 4.3: Long term profiles of the pressure-corrected normalized Scaler rate, with (black) and without (grey) the AoP correction applied [12].

4.1.2. Scaler dataset A

This dataset [13] is a raw form of the Scaler data, which is openly accessible [18]. It will be run through the same analysis method as in the GRAPES-3 publication, as both Scaler dataset A and the GRAPES-3 muon rates provide similar simple corrections.

Close to the Scaler threshold (3 ADC counts above the baseline) some detectors can behave unstably. Therefore, each individual second, detector stations with rates lower than $500 \text{ counts s}^{-1}$ were discarded. Secondly, periods with less than 97% of the entire array in operation were removed. This accounts for a loss of about 10% of the data. Afterwards, the average Scaler rate was computed for each station over the lifetime of the detector. Detector stations with a RMS of more than twice the square root of that average rate were excluded from the data. The remaining data is then corrected for pressure according to Section 4.1.1. Finally, the rates of each detector were averaged over 15-minute intervals to remove high frequency noise, that could, for example, be produced by lightning strikes.

4.1.3. Scaler dataset B

This dataset provides Scaler rates with many corrections already applied. It therefore will be used for the independent analysis as it should allow the best possible result.

It has been corrected for both the area-over-peak effect and pressure as described in Section 4.1.1. The provided rates have been averaged over 5 minutes. For a better data quality, two filters are applied to the data before analysis:

1. The three PMTs of the detector station must be working and
2. the average number of working detector stations in the array must be larger than 600 within each 5 minute interval.

After the above filters the following corrections have been applied: First, the *AoP* correction according to the fit result mentioned in Equation (4.4). Afterwards, all Scaler data $S_{\text{id}}^{\text{corr}/\text{AoP}}(t)$ more than 2.5 standard deviations from the current mean value (of the set of data of the entire array in a 5 minute window centered at time t) has been discarded. Second, for each 5 minute time bin, a centered moving average over 4 months length for the number N of working detectors and the *AoP* values has been computed. The respective mean values and standard deviations σ_N , σ_{AoP} of both quantities have been calculated and the following cuts have been applied:

4. Data description

- Periods where the current number of working detectors is outside of $\pm 1\sigma_N$ have been discarded and
- periods where the area-over-peak averaged over the entire detector $\langle AoP_{id}(t) \rangle_{vid}$ is outside $\pm 3.3\sigma_{AoP}(t)$ have also been rejected.

Finally, the pressure correction is performed. The pressure data used in this dataset [19] has been taken from the Central Laser Facility, which is located in the middle of the Surface Detector array. It has proven as the most reliable weather station and therefore has also been used as a reference for the other weather stations of the Pierre Auger Observatory. Weather data is available at different quality levels. The level chosen for this dataset is ‘L1’, which means that short gaps of $\Delta t \leq 18800$ s are interpolated linearly between the values at the edges of the gap. For this dataset a new linear fit has been performed (cf. Figure 4.4) for the *AoP*-corrected Scaler data yielding a slope of

$$m^{S,p} = -3.0 \frac{\text{‰}}{\text{hPa}} \quad (4.5)$$

so that the Scaler rate can once again be corrected via:

$$\langle S_{id}^{\text{corr}/AoP}(t) \rangle_{vid}^{\text{corr}/p} = \langle S_{id}^{\text{corr}/AoP}(t) \rangle_{vid} - m^{S,p} \cdot (p(t) - \langle p(t) \rangle_{vt}) \quad (4.6)$$

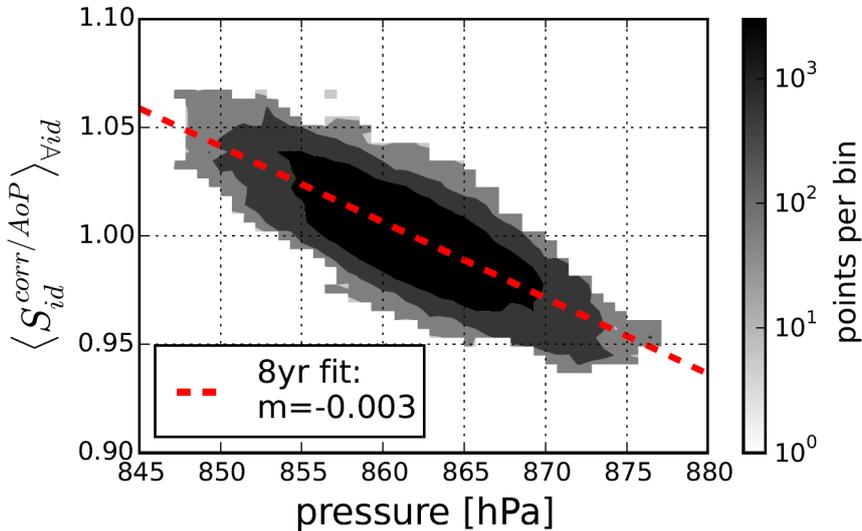


Figure 4.4: 2D histogram showing the anticorrelation between the 5-day averaged, *AoP* corrected, normalized Scaler rates and atmospheric pressure [12].

4.2. NASA OMNIWeb data

The Solar Wind velocity data used in this analysis is provided by the NASA OMNIWeb service. The Solar Wind velocity data is available for 1-minute time intervals [20]. The high resolution OMNI data are built by using data from three different satellites: ACE [21], GEOTAIL [22] and WIND [23]. The steps followed to acquire and prepare this data from the satellites can be found in detail in [24], however a brief outline can be found below.

4.2.1. Input data preparations and cleaning

The data from the three satellites each cover different portions of the Scaler data between 2006 and 2015 due to their positioning in space. IMP-8, for example, can provide Solar Wind velocity data for only 7.5 days of its 12.5 days orbit.

In addition to the different data availability, the three satellites also provide different time averages of the data that have to be merged:

- ACE: 64-second averages
- GEOTAIL: 48-second averages
- WIND: 92-second averages

There are both statistical and systematic differences between the different datasets. These can be caused by having different time gaps in their averaged data, spatial gradients in the parameters being measured combined with offsets of the spacecraft locations relative to the flow direction of the Solar Wind, wrong time-shifts or different data processing methods (fits or derivation from distribution functions). The systematic differences change rather slowly, so these effects have been taken care of by using the hourly averaged data. The WIND dataset has been chosen as the benchmark because its parameters' uncertainties have very well been tested by comparing proton and alpha particle densities [25]. The other two datasets have then been fitted linearly to WIND data [26].

Afterwards the data has been run through a despiking algorithm: For any given datapoint x_i the mean value $\langle x \rangle$ and root mean square $\sigma(x)$ have been calculated for the point x_i itself and its two predecessors x_{i-1}, x_{i-2} and two followers x_{i+1}, x_{i+2} . These values must lay within a 60 minute range of the original datapoint, otherwise they have been removed to eliminate huge gaps right away. In a first test, datapoints fulfilling $|x - \langle x \rangle| > 4 \cdot \sigma(x)$ have been rejected. Secondly, datapoints require $|x - \langle x \rangle| > 0.1 \cdot \langle x \rangle$ to remain in the final dataset.

4.2.2. Time shifting

The goal of time-shifting the Solar Wind velocity data is to account for the time difference between the location of observation and a point close to the Earth's magnetosphere, called the bow shock nose. It is assumed that Solar Wind variations are organized in a series of flat plane phase fronts convecting with the Solar Wind velocity \vec{v}_{SW} . Hence, the time-shift of two observations at points $P_1(t_1, \vec{r}_1)$ and $P_2(t_2, \vec{r}_2)$ can be expressed as

$$\Delta t = \frac{\vec{n} \cdot (\vec{r}_2 - \vec{r}_1)}{\vec{n} \cdot \vec{v}_{\text{SW}}} \quad (4.7)$$

where \vec{n} is the so-called *phase front normal*. The determination of the phase front normals is a difficult process involving minimum variance analysis and certain modifications of it, which shall not be described in detail here but can be found in [27].

4.2.3. Building 1-minute averages

The output data is tagged at the start of every minute. As the input data averages cover different time spans, a 1 minute average of the output data can contain data from up to 3 separate input data time averages. Because of this, a weighted average is needed. This problem is depicted in Figure 4.5. For every datapoint contributing to one output minute the fraction of the current output minute covered by that input datapoint is calculated and applied as a weight to that input period and the weighted average of all contributing datapoints is calculated. Similarly, a weighted variance is computed.

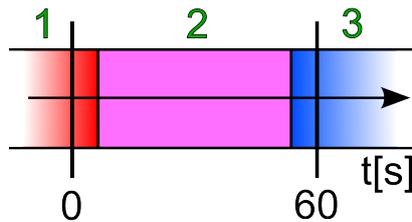


Figure 4.5: Example for averaging GEOTAIL data (48 second averages): The desired output average spans from 0s to 60s. Assuming that input sample 2 is centered in that minute, sample 2 contributes to a fraction of $\frac{48}{60}$, while 1 and 3 contribute to $\frac{6}{60}$ each.

5. Simple analysis with few corrections

The main purpose of this study is to either verify or disprove the GRAPES-3 result [1] of the anticorrelation of the Solar Wind velocity variation with cosmic ray intensity. In this chapter their experimental setup and methods of data treatment will be explained. Afterwards, the analysis of GRAPES-3 is applied to the Scaler data of the Pierre Auger Observatory. Here, the Auger Scaler dataset A (cf. Section 4.1.2) is used, as it provides similar basic cuts for data quality and its rather low resolution of 15 minutes is sufficient. The results are compared to those of GRAPES-3.

5.1. GRAPES-3 experiment

The GRAPES-3 (= **G**amma **R**ay **A**stronomy **P**eV **E**nergie**S** phase-**3**) experiment is located at Ooty (11.4°N, 76.7°E, 2200 m a.s.l.) in India. It has two detectors to measure the flux of cosmic rays: Firstly, a dense air shower array of 400 scintillation detectors with an area of 1 m² each at an 8 m spacing, whose task is to measure the energy of primary cosmic rays in an energy range of 10 TeV to 100 PeV. Secondly, a large muon detector with an area of 560 m² to determine the composition and the energy spectra of the muon content of cosmic ray showers. The latter is also used to measure the modulation of cosmic rays by solar activity and provides all data for the anticorrelation analysis. The detector consists of 16 modules with an area of 35 m² each, which each contain 232 proportional counters arranged in four layers of 58 counters each. These are separated by a concrete layer, resulting in an energy threshold of 1 GeV for vertical muons.

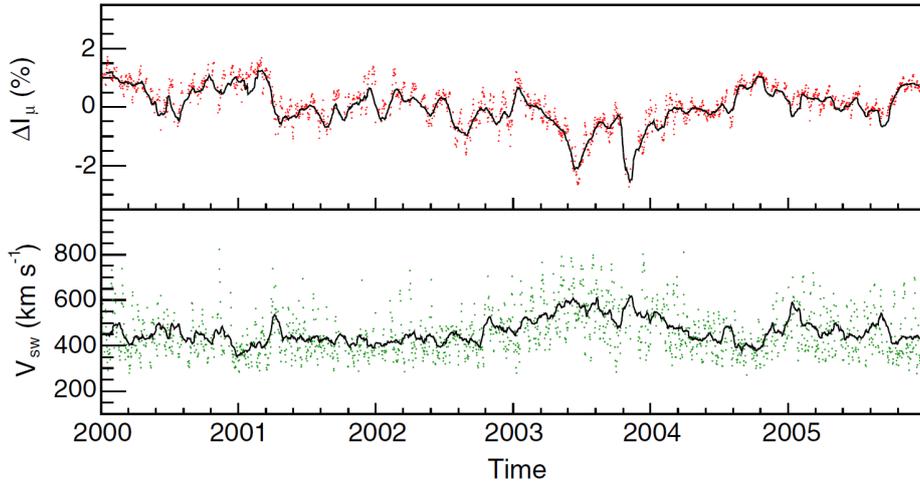
Every 10 s the number of incident muons is recorded and hourly averaged data of six complete years from 2000 to 2005 is used in their analysis.

5.2. GRAPES-3 data treatment

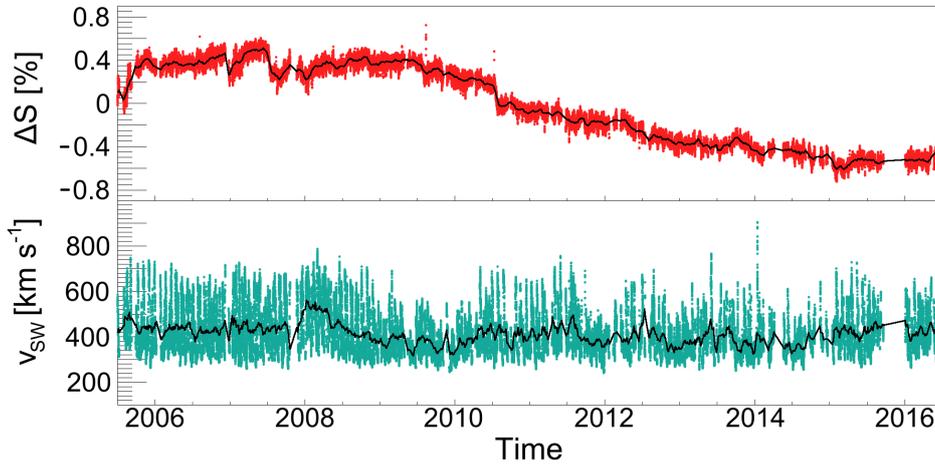
The effect of instrument problems resulting in e.g. gaps and gain variation in the muon rate is first taken care of by rejecting any hourly data deviating more than 10 times the RMS from the overall average. Afterwards, the mean and RMS are recalculated and data

with a deviation greater than 5 times the RMS is thrown out. A pressure correction is then applied to the data. The hourly muon rates are then converted to fractional changes from the six-year mean of the remaining data.

To remove the period modulation due to the solar rotation, a centered moving average is used. For every hourly intensity the centered moving average of 26 days is calculated which afterwards is averaged to daily values. In Figure 5.1(a) the daily data for muon intensity and the Solar Wind velocity as well as the values of their centered moving averages are shown.



(a) GRAPES-3 data



(b) Scaler rate of the Pierre Auger Observatory

Figure 5.1: Daily variation in (a) muon intensity ΔI_μ as measured by GRAPES-3 (relative to respective 6-year mean) [1] and (b) Scaler rate ΔS (relative to respective 10-year mean) and Solar Wind velocity v_{SW} . The solid lines show the corresponding 27-day centered moving average.

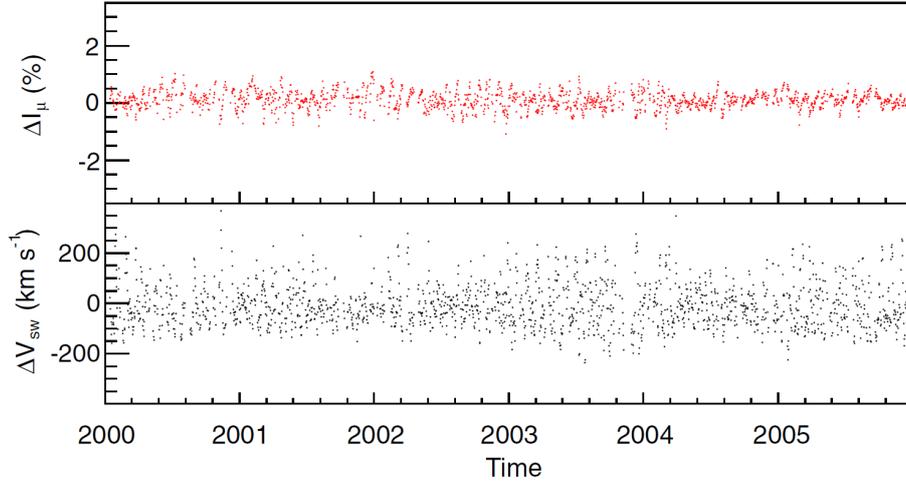
After that, the combined daily value of the centered moving average is subtracted from the data. Finally, Forbush decreases are removed. This is done using data from the Kiel neutron monitor (see [28]): for each day i the neutron intensity I_i and the mean intensity of the previous three days $(i - 1)$, $(i - 2)$, $(i - 3)$ called $I3d_i$ is computed. If the relative decrease in intensity $\frac{I3d_i - I_i}{I3d_i}$ is greater than 2%, then data from day i is rejected. The days following day i are rejected until the intensity again surpasses the value of $0.3 \cdot I_i + 0.7 \cdot I3d_i$, which corresponds to a 70% complete recovery from the Forbush decrease. As only ‘complete’ days with data for all 24 hour intensities for all three data sets (muon, Solar Wind velocity and neutron monitor data) are considered, only 2117 days of initially 2192 days survive. The resulting data is shown in Figure 5.2(a). It displays a flat profile as periodic and transient effects have been removed.

Similar methods of data cleaning have been applied to Scaler dataset A (see Section 4.1.2). The RMS cuts are not necessary because the dataset is already filtered to eliminate bad data due to malfunctioning as was described in Section 4.1.2. From the high-resolution 1-minute averaged OMNIWeb data 15-minute averages have been computed to be matched to the 15-minute Scaler data. Demanding ‘complete’ days is impossible with this Solar Wind velocity dataset, therefore 15-minute intervals are demanded to have at least 8 minute-long entries, otherwise they will be rejected. Unfortunately, the Scaler data has a lot of gaps, so that when averaging Solar Wind velocity and Scaler data for daily values both are required to have data for at least 50% of each day, otherwise data from that day is rejected completely. Running means are computed and subtracted and Forbush decreases are filtered out using data from the McMurdo neutron monitor [29], because the Kiel neutron monitor was lacking data for the time span of the Scaler data (2006-2015). After the entire cleaning process there are 3116 days of data left.

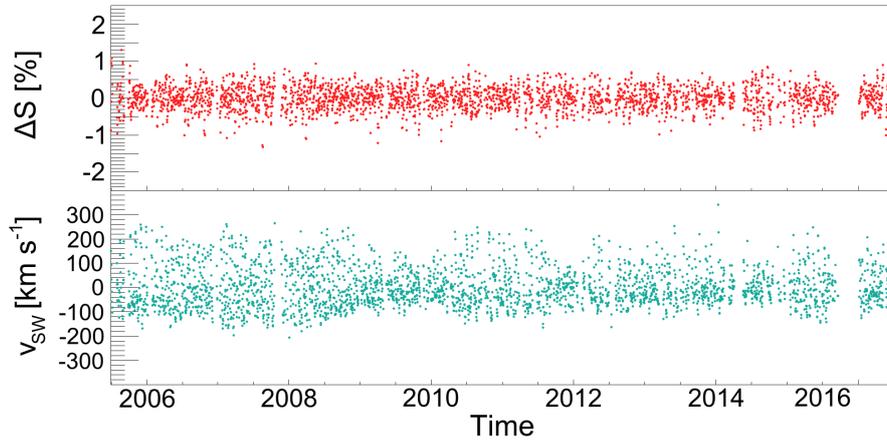
The resulting analogous plots can be found in Figure 5.1(b) and Figure 5.2(b). Comparing both plots in Figure 5.1 one can see that the centered moving averages resemble the data well. Looking at Figure 5.2 the variation of the muon / Scaler rate is flat and in the range of about 1% for both experiments. In addition to that, the variation in the solar wind is similar at about $200 \frac{\text{km}}{\text{s}}$ for both experiments.

5.3. Comparison of GRAPES-3 muon intensity and Scaler data results

After applying all cuts mentioned above, the correlation of CR intensity and Solar Wind velocity is calculated. In Figure 5.3 a scatterplot comparison of both experiments is displayed. Although the general trend of the data seems comparable, the fluctuation of data seems to be a bit higher for the Scaler data.



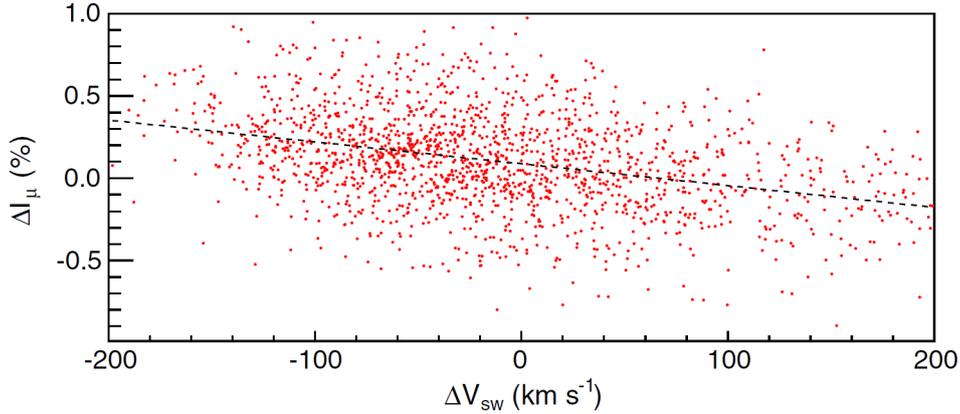
(a) GRAPES-3 data



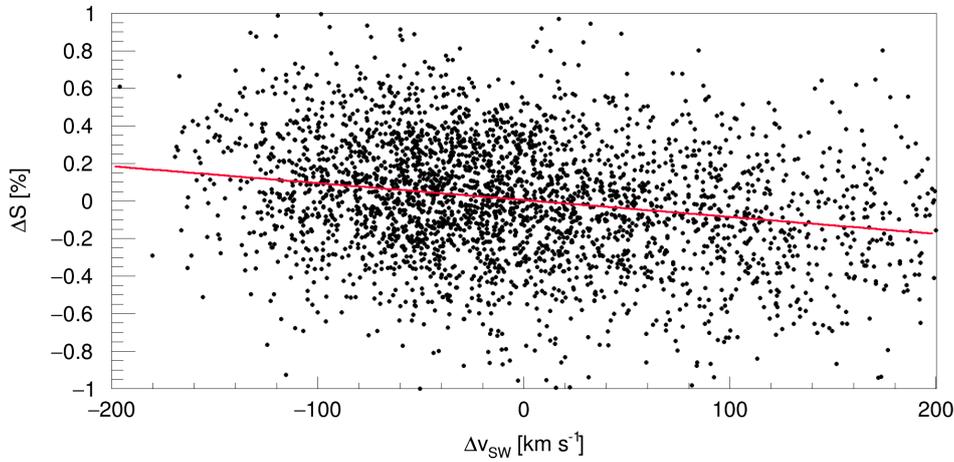
(b) Scaler rate of the Pierre Auger Observatory

Figure 5.2: Daily variation after high pass filter: (a) muon intensity ΔI_μ as measured by GRAPES-3 (relative to respective 6-year mean) [1] and (b) Scaler rate ΔS (relative to respective 10-year mean) and Solar Wind velocity deviation Δv_{sw} (from the respective mean value).

Afterwards, the daily data of GRAPES-3 was combined into 41 bins, each containing data from 43 d sorted in order of ascending v_{sw} . Analogously, the Scaler data is combined into 73 bins of 43 days each to have a comparable amount of final datapoints (the last bin only has 20 entries). The final plots in Figure 5.4 feature linear fits, which reasonably describe the data for both experiments. The results for the slope parameter listed below.



(a) GRAPES-3 data



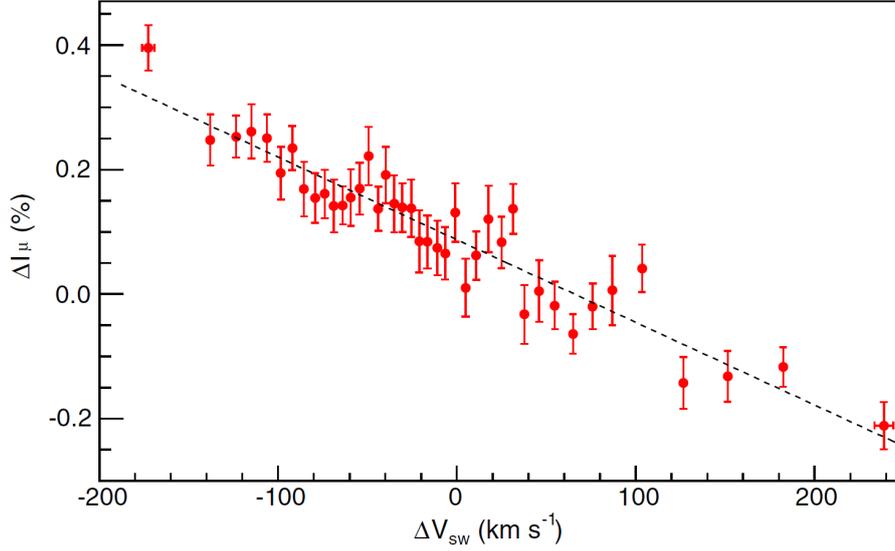
(b) Scaler rate of the Pierre Auger Observatory

Figure 5.3: Correlation of (a) muon intensity variation ΔI_μ as measured by GRAPES-3 [1] and (b) Scaler rate ΔS with Solar Wind velocity Δv_{SW} including linear fits.

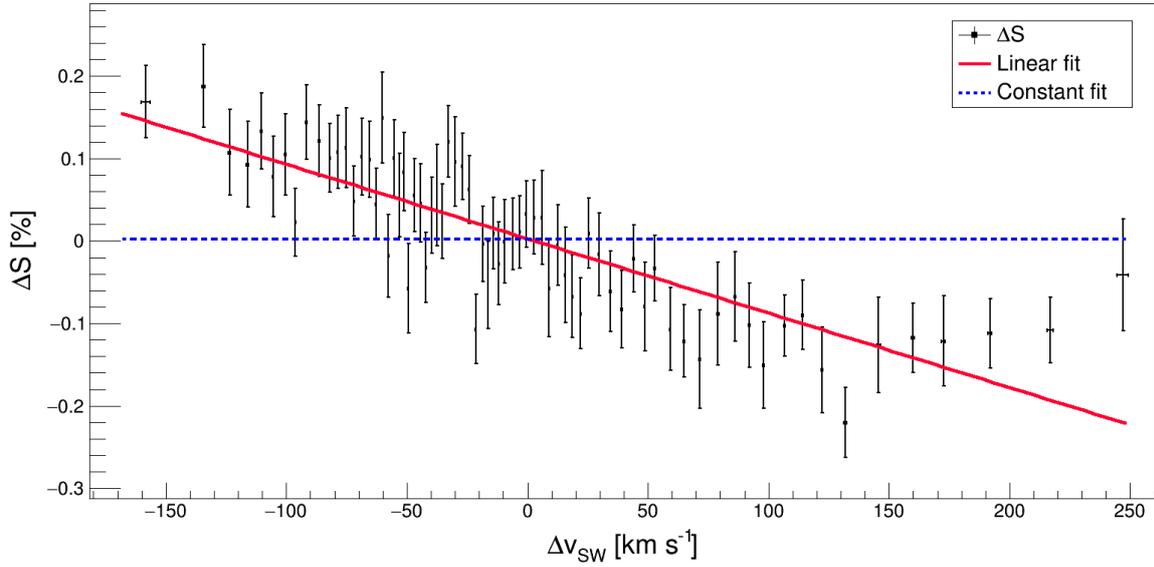
$$\begin{aligned}
 m_{\text{corr}}^{\text{GRAPES}} &= [-13.3 \pm 0.7 \text{ (stat.)}] \frac{10^{-4}\%}{\text{km} \cdot \text{s}^{-1}} \\
 m_{\text{corr}}^{\text{Scaler,A}} &= [-9.0 \pm 0.6 \text{ (stat.)}] \frac{10^{-4}\%}{\text{km} \cdot \text{s}^{-1}}
 \end{aligned}
 \tag{5.1}$$

One can see that the effect is very small: For a typical change of Solar Wind velocity of about $200 \frac{\text{km}}{\text{s}}$ a relative deviation of $\sim 0.2\%$ from the mean value of muon / Scaler rate is expected. Both results significantly differ from the null-hypothesis of a constant

function, which would be expected if there is no correlation at all. However, both results are not compatible within the error ranges of the fit parameter.



(a) GRAPES-3 data



(b) Scaler rate of the Pierre Auger Observatory

Figure 5.4: (a) Muon intensity variation ΔI_{μ} as measured by GRAPES-3 [1] and (b) Scaler rate correlated with Solar Wind velocity Δv_{sw} after binning into 43-day intervals.

The linear fit (cf. Figure 5.4(b)) of the Scaler data has a reduced χ^2 of 1.17 which shows, that the uncertainties are estimated correctly and the model describes the data well. Additionally, a constant function is fit to the data, but this model can be ruled out due to its high reduced χ^2 of 3.87. The different results for the slope can have two reasons: Either there is a physical reason for them two be different, e.g. the anticorrelation slope could depend on the location on Earth, or one of the results is wrong. To verify the result using Scaler data and to potentially lower the uncertainty on the slope parameter, an independent analysis of the Scaler data with different methods and more systematic cuts will be presented in the next chapter.

6. Independent analysis of the anticorrelation

Although the anticorrelation could already be shown with Scaler dataset A, the uncertainty on the fit result is larger than expected. With Scaler data spanning 10 years (4 years more than GRAPES-3 had), it should be possible to lower the uncertainty on the anticorrelation result by conducting an independent analysis and by using Scaler dataset B, which has more recently studied systematical corrections like the AoP -correction already applied. The data is smoothed to filter small scale statistical fluctuations while not reducing the signal on a larger scale.

6.1. Determining a time-window

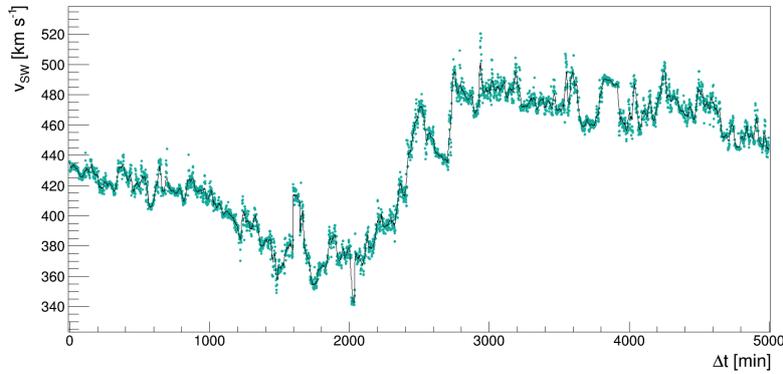
For the GRAPES-3 analysis the Solar Wind velocity and muon rate time window was set to 60 min without further explanation and later averaged again over 24 h. This section is dedicated to the search for the best fitting time-window δt for both Scaler and Solar Wind velocity data.

First, 10 random time-frames of data from 10 different years i with $i = 2006, \dots, 2015$ are picked at random. The spread in years is needed to avoid a dependence on the solar magnetic cycle. The length of the time-frame is 27 days to minimize the effects from solar rotation. This means, that one data sample, in theory, contains 38880 Solar Wind velocity data entries (1-minute averaged data) and 7776 Scaler data entries (5-minute averaged data). In practice, there are less entries as both datasets occasionally have small gaps. For every random sample a running mean is computed for every datapoint (so for every minute or every five minutes respectively) using a δt of varying length (5 min to 120 min). It will be tested for which length of δt the running mean fits best to the data. This result for δt will afterwards be used as the averaging time window for both Scaler and Solar Wind velocity data.

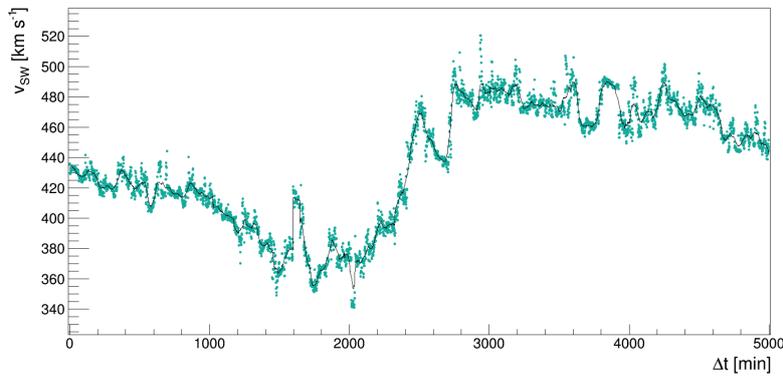
An example comparing running means of width $\delta t = 10$ min, 30 min, 120 min for the Solar Wind velocity data is shown in Figure 6.1. However, only the first 5000 min of the full 38880 min from the 27-day period are shown to make the details visible. The quality of the running mean curve is characterized by two properties:

6. Independent analysis of the anticorrelation

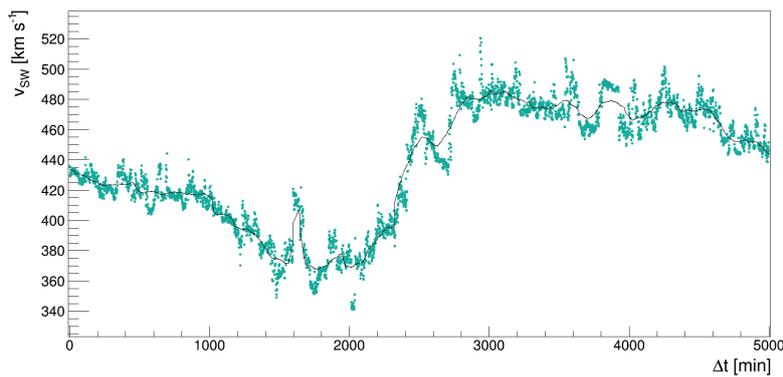
1. How well the running mean resembles the original data and
2. how smooth (i.e. small fluctuations and noise filtered out) the curve is.



(a) $\delta t = 10$ min



(b) $\delta t = 30$ min



(c) $\delta t = 120$ min

Figure 6.1: First 5000 min of a random 27 d sample of Solar Wind velocity data starting on 22.03.2009 at 08:04. 1-minute averages provided by OMNIWeb (green dots) and running mean with lengths of 10, 30, and 120 minutes (black lines).

For the first example one can see that the 10-minute running mean curve (see Figure 6.1(a)) is very close to the original data, but there still is some noise included. Using a 120-minute running mean (see Figure 6.1(c)), some features of the original data – variations in Solar Wind velocity of up to $30 \frac{\text{km}}{\text{s}}$ – are lost. This information loss is especially visible in the period between 500 min and 1000 min. For the Solar Wind data, a 30-minute running mean δt as in Figure 6.1(b) seems to be best: It filters the noise on small time scales while preserving the larger scale features. Additional plots for different values of δt and for different years can be found in Appendix A.1.

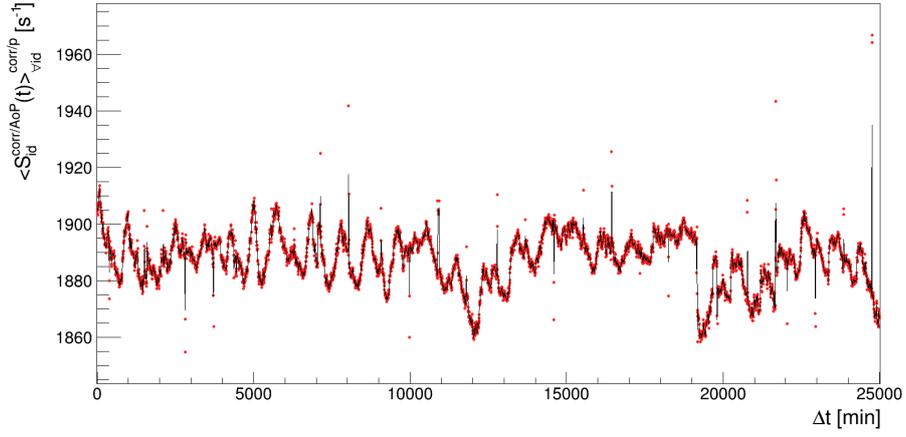
For a second example, the running mean of Scaler data with a length of $\delta t = 10$ min, 30 min and 120 min is presented in Figure 6.2. Some of the datapoints are far from the running mean and cause spikes in the running mean data that do not vanish until δt is increased to at least 90 min. Additional plots for different values of δt and for different years can be found in Appendix A.1.

To quantify the two characteristics mentioned above (resembling the original data and smoothness), the following quantities are defined:

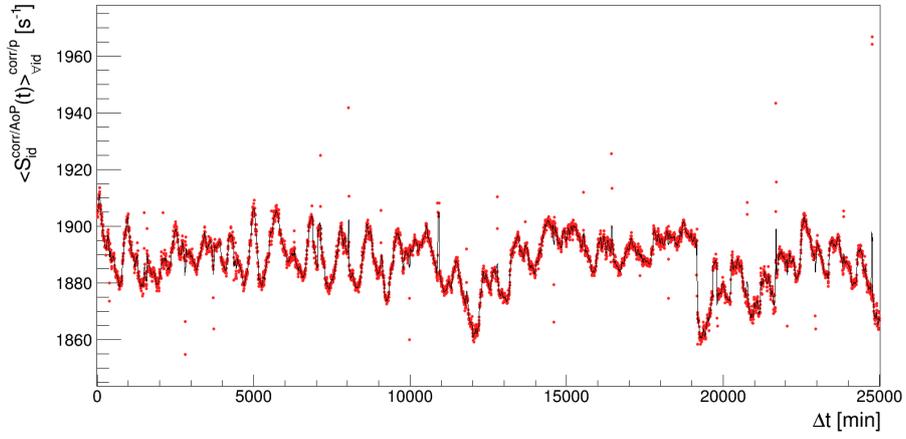
1. $|\overline{\Delta v}|$: The 27-day average of the difference between the running mean and the original Solar Wind velocity data.
2. $|\overline{\Delta S}|$: The 27-day average of the difference between the running mean and the original Scaler data.
3. $\overline{\sigma_{v,dt}}$: The average standard deviation of the running mean Solar Wind velocity data when binned into time-windows of length dt .
4. $\overline{\sigma_{S,dt}}$: The average standard deviation of the running mean Scaler data when binned into time-windows of length dt .

In addition to the length δt of the running mean, the time-window dt for computing a standard deviation is introduced. Clearly, the smaller δt is, the lower the $|\overline{\Delta v}|$ and the $|\overline{\Delta S}|$ of the running mean data. In the limit $\delta t \rightarrow 0$ the deviations would all be zero. On the contrary, the smaller the length of δt , the larger $\overline{\sigma_{v,dt}}$ and $\overline{\sigma_{S,dt}}$ tend to be. With shorter running mean lengths the running mean data inherits the random fluctuations of the 1-minute or 5-minute averaged original data. Finding the right time-window length dt for the running mean standard deviations is difficult, because it is related to finding the right running mean length δt . Hence, time-windows of length dt varying from 15 min to 120 min are used. For each time-window dt and each randomly chosen datasample the dependence of $|\overline{\Delta v}|$ and $\overline{\sigma_{v,dt}}$ (and $|\overline{\Delta S}|$ and $\overline{\sigma_{S,dt}}$ respectively) on δt is tracked. Both graphs are fit with logarithmic function $a \cdot \ln(\delta t - b) + c$ which is chosen, because it describes the data well.

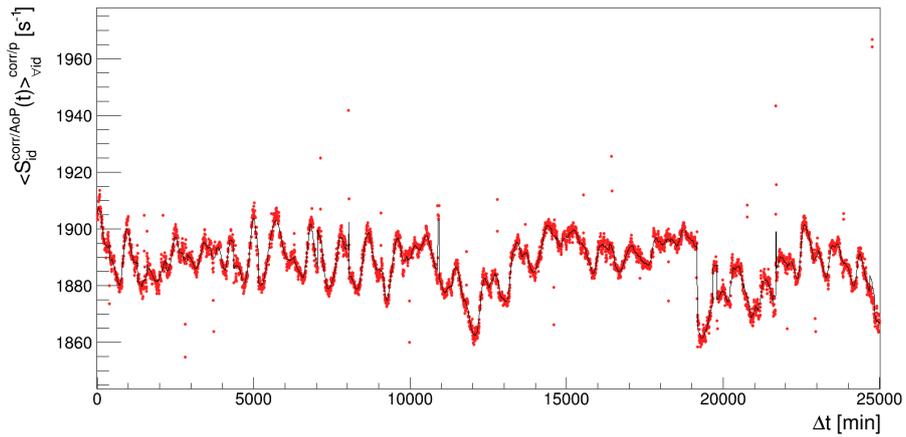
6. Independent analysis of the anticorrelation



(a) $\delta t = 10$ min



(b) $\delta t = 30$ min



(c) $\delta t = 120$ min

Figure 6.2: Random 27 d sample of Scaler data starting on 26.02.2011 at 13:35 with the original 5-minute averages (red dots) and the running mean with lengths of 10, 30 and 120 minutes (black lines).

The intercept point of both fits – i.e. the point where the average difference of running mean data from the original data is the same as the average standard deviation of the running mean data themselves – provides a good choice of δt for each dataset. An example of this procedure carried out on the Solar Wind velocity data is illustrated in Figure 6.3.

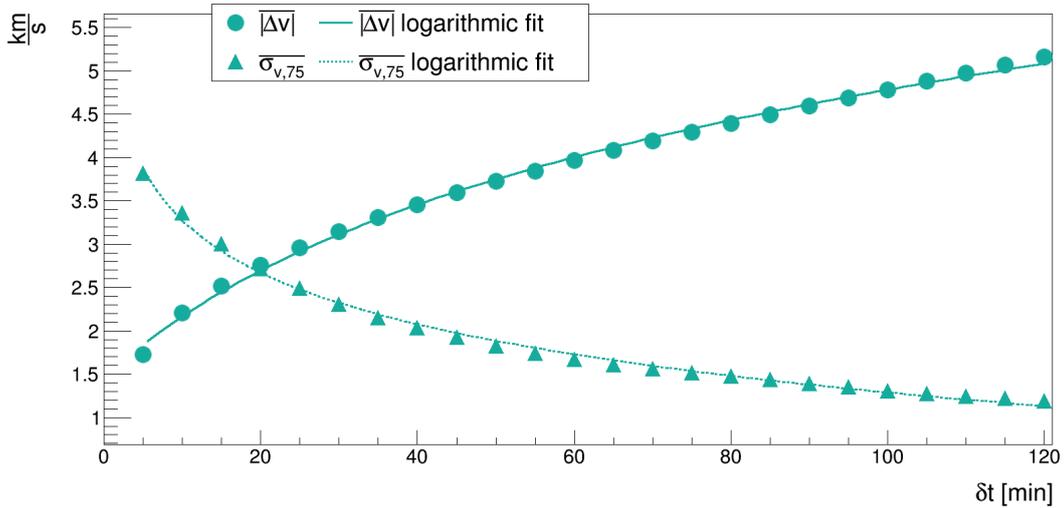


Figure 6.3: Dependence of $|\overline{\Delta v}|$ (blue dots) and $\overline{\sigma_{v,75}}$ (blue triangles) on the length of δt for a random Solar Wind data sample taken from 2009.

For $dt = 15$ min, the starting value for $|\overline{\Delta v}|$ was already higher than that of $\overline{\sigma_{v,dt}}$. This data is not included in the study, as the extrapolations of the fits with $\delta t < 5$ min cannot be trusted to correctly estimate the time. All other results for the intercepts of Solar Wind velocity data are listed in Table 6.1.

The mean value of the intercept of both graphs is dependent on the length of the time-window that is used to compute $\overline{\sigma_{v,dt}}$ and it seems to grow linearly with the length of that interval dt . However, by looking at the data, any running mean length $\delta t < 15$ min seems too low and is too close to the Scaler data time interval of 5 min to provide a mean value with a good statistical uncertainty.

An example for the two fits using Scaler data is shown in Figure 6.4, using a random data sample from 2011 and $dt = 75$ min. In Table 6.2 the intercepts for varying values of dt are listed. They show the same behaviour as before: With increasing dt the intercept also increases. The mean intercept points of Scaler data for a fixed dt are always higher than their respective counterparts for the Solar Wind velocity data. The difference between both mean intercept values also increases for increasing dt . Additional plots for different values of δt and for different years can be found in Appendix A.1.

6. Independent analysis of the anticorrelation

Table 6.1: Solar wind data: intercepts of the $\overline{|\Delta v|}$ and $\overline{\sigma_{v,dt}}$ fits for all datasamples depending on dt . Averages for a fixed dt in green highlighted cells.

$dt = 30$ min		$dt = 45$ min		$dt = 60$ min		$dt = 75$ min	
sample year	intercept [min]						
2006	6.9	2006	10.7	2006	14.9	2006	19.0
2007	7.0	2007	10.1	2007	14.5	2007	17.3
2008	6.7	2008	10.9	2008	15.6	2008	19.9
2009	6.6	2009	10.6	2009	15.1	2009	19.7
2010	5.6	2010	8.4	2010	11.7	2010	15.3
2011	5.1	2011	7.6	2011	10.6	2011	14.1
2012	4.5	2012	5.4	2012	7.7	2012	10.3
2013	6.6	2013	10.3	2013	15.2	2013	19.4
2014	7.0	2014	10.5	2014	13.8	2014	17.4
2015	5.9	2015	8.3	2015	11.1	2015	14.0
	6.2		9.3		13.0		16.6

$dt = 90$ min		$dt = 105$ min		$dt = 120$ min	
sample year	intercept [min]	sample year	intercept [min]	sample year	intercept [min]
2006	24.3	2006	28.4	2006	33.4
2007	21.5	2007	25.8	2007	29.6
2008	24.0	2008	29.9	2008	34.5
2009	23.4	2009	28.3	2009	34.2
2010	19.9	2010	23.6	2010	26.6
2011	17.2	2011	20.0	2011	24.6
2012	13.4	2012	16.3	2012	18.9
2013	25.7	2013	28.4	2013	38.4
2014	21.5	2014	24.3	2014	28.7
2015	17.8	2015	20.7	2015	23.5
	20.9		24.6		29.2

6. Independent analysis of the anticorrelation

Table 6.2: Auger Scaler data: intercepts of the $|\overline{\Delta S}|$ and $\overline{\sigma_{S,dt}}$ fits for all datasamples depending on dt . Averages for a fixed dt in green highlighted cells.

$dt = 30$ s		$dt = 45$ s		$dt = 60$ s		$dt = 75$ s	
sample year	intercept [min]						
2006	13.3	2006	18.2	2006	23.4	2006	30.8
2007	10.8	2007	14.4	2007	19.1	2007	22.7
2008	10.7	2008	13.5	2008	20.1	2008	27.2
2009	8.1	2009	10.6	2009	13.3	2009	15.2
2010	10.5	2010	13.7	2010	16.8	2010	21.4
2011	12.3	2011	16.6	2011	20.7	2011	23.8
2012	9.2	2012	11.7	2012	14.8	2012	17.7
2013	14.3	2013	17.8	2013	23.8	2013	27.8
2014	15.8	2014	21.8	2014	28.9	2014	32.3
2015	12.9	2015	19.1	2015	27.4	2015	36
	11.8		15.7		20.8		25.5

$dt = 90$ s		$dt = 105$ s		$dt = 120$ s	
sample year	intercept [min]	sample year	intercept [min]	sample year	intercept [min]
2006	36.2	2006	45.6	2006	52.7
2007	37.5	2007	33.2	2007	40.8
2008	31.3	2008	37.4	2008	45.6
2009	18.4	2009	22.4	2009	25.0
2010	24.8	2010	27.3	2010	33.9
2011	32.0	2011	37.2	2011	41.4
2012	18.9	2012	22.3	2012	24.9
2013	29.9	2013	36.1	2013	37.8
2014	37.8	2014	42.4	2014	52.2
2015	45.4	2015	59.7	2015	72.2
	31.2		36.4		42.6

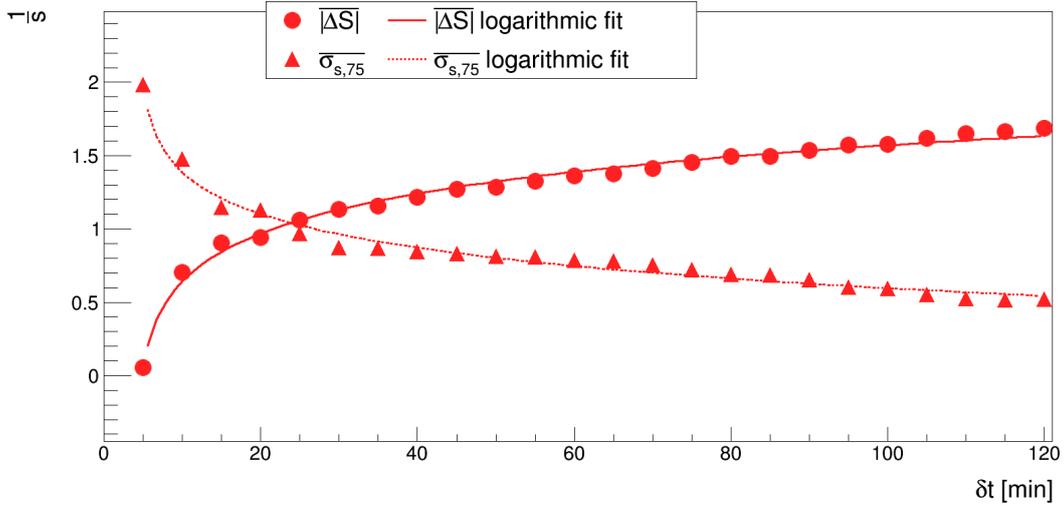


Figure 6.4: Dependence of $|\Delta S|$ (red dots) and $\overline{\sigma_{S,75}}$ (red triangles) on the length of δt for a random Scaler data sample taken from 2011.

These results make it difficult to give a precise answer for the averaging time-window δt . However, a minimal averaging time-window of $\delta t = 30$ min can be demanded of both the Scaler and Solar Wind data. The δt will remain a parameter for the entire analysis, although mainly the results for $\delta t = 30$ min will be discussed. To study the systematic effect of the choice of δt , the analysis will be performed with the set of values (30, 45, 60, 75, 90) minutes for δt which is balanced distributed around the GRAPES-3 value of 60 min.

6.2. Scaler data treatment

There are some phenomena that only affect the Scaler data and have to be corrected for before starting the correlation analysis. Their origin and effect on the data are discussed in the following sections.

6.2.1. Removing transient effects

To remove periods that are affected by Forbush decreases from the data, the same algorithm as explained in Section 5.2 is applied. Furthermore, lightning periods [30] and so-called ‘bad periods’ [31] are discarded as they pose major acquisition instabilities of the surface detector.

6. Independent analysis of the anticorrelation

The percentage of data lost per year due to each of the different criteria is listed in Table 6.3. As an example the gaps for 2015 are visualized in Figure 6.5 (data gaps for the other years can be found in Appendix A.2). Forbush decreases often last for 10 days or more, while bad periods or lightning periods mostly last only for few hours. One notices that bad periods and lightning periods often overlap, however they do not necessarily depend on each other. The stacked histogram in Figure 6.6 shows the relative losses due to Forbush decreases, lightning periods outside of Forbush decrease times and bad periods outside of Forbush decreases or lightning periods (also visible as gray highlighted columns in Table 6.3).

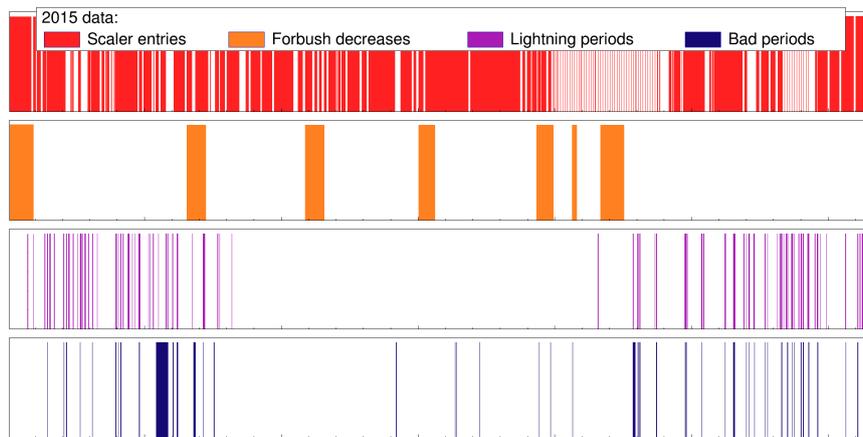


Figure 6.5: Scaler data availability (first row, red) and periods that are removed from the data due to Forbush decreases (second row, orange), lightning periods (third row, purple) and ‘bad periods’ (fourth row, dark blue) for 2015.

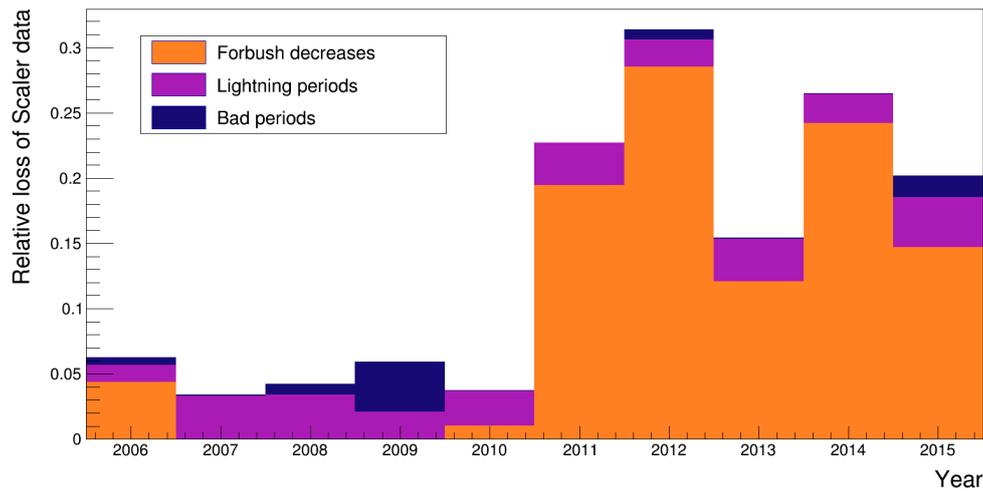


Figure 6.6: Relative loss of Scaler data for the years 2006-2015 for Forbush decreases, lightning periods and bad periods applied subsequently.

Table 6.3: Fractions of Scaler data influenced by Forbush decreases (FD), lightning periods (LP) and ‘bad periods’ (BP) for years 2006-2015. With ‘LP \ FD’ all lightning periods not removed due as Forbush decreases are denoted. Similarly ‘BP \ (FD \cup LP)’ stands for all bad periods not removed due to Forbush decreases or lightning periods.

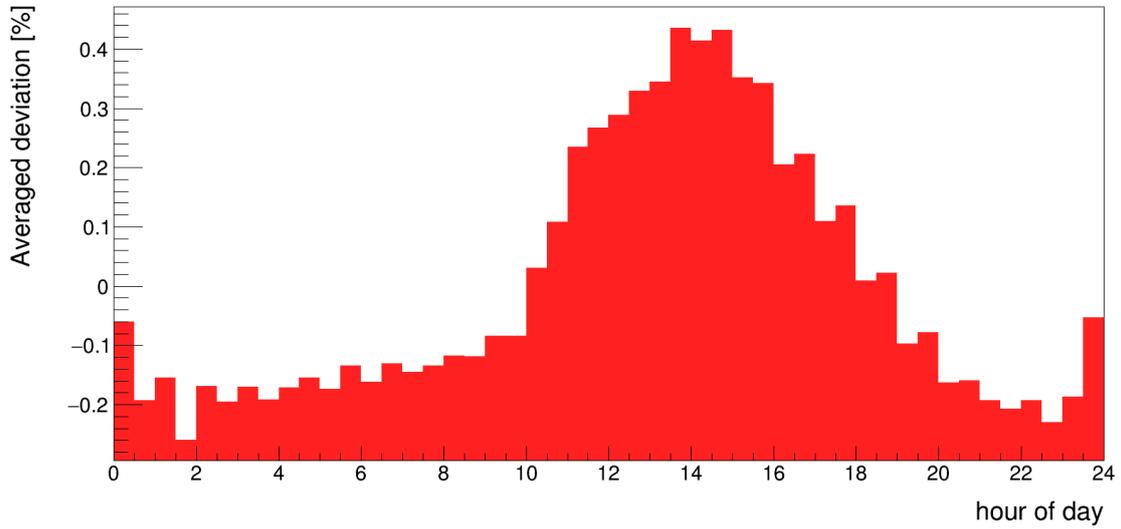
year	FD [%]	LP [%]	LP \ FD [%]	BP [%]	BP \ (FD \cup LP) [%]
2006	4.37	1.50	1.31	0.76	0.52
2007	0	3.33	3.33	0.37	0.06
2008	0	3.39	3.39	1.19	0.85
2009	0	2.08	2.08	3.97	3.82
2010	1.04	2.68	2.68	0.13	0.01
2011	19.45	3.51	3.25	0.40	0.01
2012	28.51	3.56	2.08	1.44	0.78
2013	12.07	3.83	3.26	0.46	0.10
2014	24.21	3.79	2.21	0.89	0.07
2015	14.71	4.24	3.85	2.87	1.65

6.2.2. Removing daily patterns

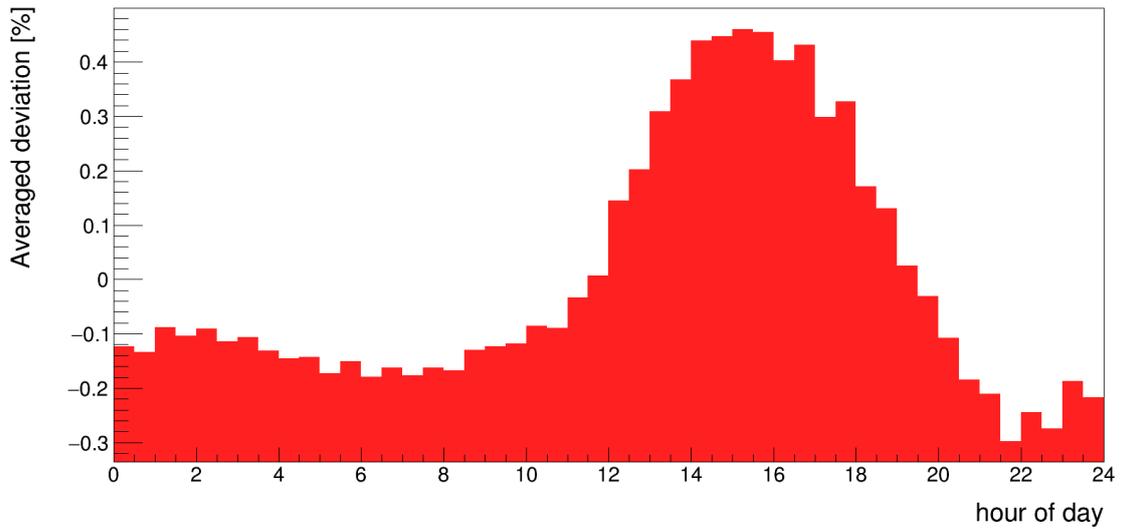
The Scaler rate of the surface detector stations has been observed to feature a daily oscillation [32]. This daily pattern also changes depending on the season. Therefore, the relative deviation from the mean daily Scaler rate in a certain month of the year as a function of UTC hour of day has been averaged over the entire dataset and it is binned with the time-window δt as bin width. An example using a time-window of 30 min is shown in Figure 6.7. It displays the daily oscillation in January and July corresponding to summer and winter in Argentina (data for other months can be found in Appendix A.2). The overall variations are within $\pm 0.4\%$. While this is small, it is still on a similar order as the anticorrelation with Solar Wind velocity and would therefore produce an unwanted bias. The peak of the distribution occurs at about 15:00 UTC, which is 12:00 Argentina time, showing correlation with the position of the sun in the sky. In January, the averaged deviation rises above 0 at about 10:00 UTC and stays positive until 19:00 UTC. In contrast to that, the average deviation in July is only positive between 12:00 UTC and 19:00 UTC. This can be explained by seasonal effects: In the Argentinean summer the sun is up longer than in the winter which widens the peak in the daily oscillation. To correct for this daily oscillation, Scaler rates of time bin i will be multiplied

6. Independent analysis of the anticorrelation

with $(1 - S_{\text{daily},i})$, where $S_{\text{daily},i}$ is the bin content of the i -th bin of the average deviation histograms mentioned above.



(a) January



(b) July

Figure 6.7: Average deviations from the daily mean Scaler rates for a chosen time-window of 30 min for (a) January and (b) July.

6.3. Minimal required entries

Finding the right time-window for the analysis has been difficult, but, as mentioned earlier, it is kept between 30 min and 90 min. There will be no further averaging applied unlike in the GRAPES-3 analysis, where daily averages and later on even multi-day averages have been formed. An absolute minimum of 2 entries per time bin are required since otherwise a sample standard deviation could not be defined.

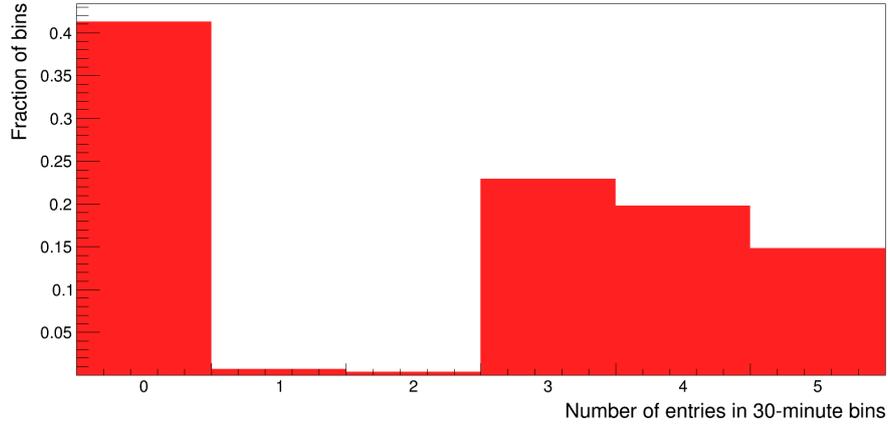
The main motivation for the independent analysis is to use the larger statistics of the Scaler data in the anticorrelation fit and obtain a result with a higher precision than the GRAPES-3 result. This should be possible as the Scaler dataset covers a larger time span than that available to GRAPES-3. To lower the uncertainty on the mean value of the averaged data, the minimum number of data entries per time bin can be raised well above the bare minimum of 2. Enforcing stricter requirements like this increases the loss of data, as for most time bins data are not available for the entire length of the bin due to the various cuts applied to the data which have been performed before averaging over δt .

To determine the optimal trade-off between data loss and improving the statistical uncertainty, the distribution of entries in a given time binning is analyzed. In Figure 6.8 the entry distributions of the Scaler data are shown. In this example the distribution of the number of entries for time bins of 30 min (see Figure 6.8(a)) will be used:

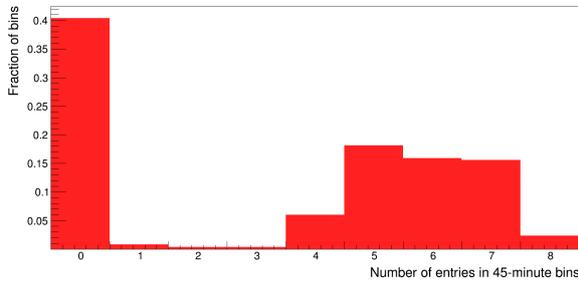
Due to the various cuts that are applied to the data, $\sim 41\%$ of all 30-minute bins of the years 2006-2015 do not have data. With increasing averaging length δt , the number of empty bins decreases to $\sim 38\%$ at $\delta t = 90$ min. With this time-window length there are very few 30-minute intervals with only 1 or 2 entries, only making up about 1% of the full 10 years. Therefore, a minimum of 3 entries is set as a requirement to include data for the 30-minute averages. Having few time intervals with few entries until a certain threshold is also true for other choices of δt , as can be seen in Figures 6.8(b) to 6.8(e). Therefore, to provide a lower uncertainty of the mean value, which decreases as the square root of the number of entries, these thresholds are set as a requirement for the minimal number of entries.

Looking at the corresponding plots for Solar Wind velocity data in Figure 6.9, there are two striking differences: First, the number of empty bins is below 5% for all δt s and there is no threshold behaviour. Second, the percentage of bins that have one or more data points increases with the number of entries until a maximum is reached. The position of the maximum is between 90% and 100% of the respective maximal number of entries. That means, if one time-window of length δt is picked at random, the probability that there is data available for about 90% of the chosen time interval, is maximal among all other portions of data available. To avoid bias, for every δt , the threshold values of

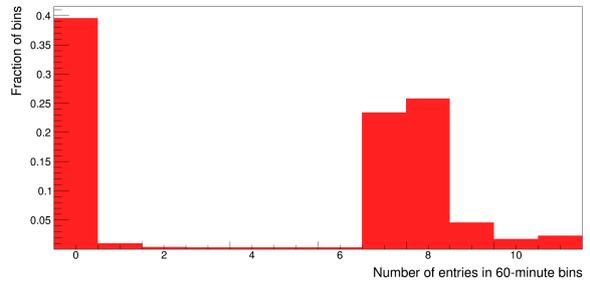
6. Independent analysis of the anticorrelation



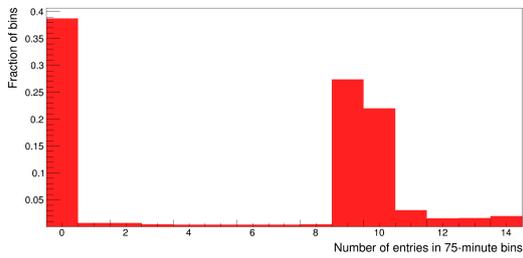
(a) $\delta t = 30$ min



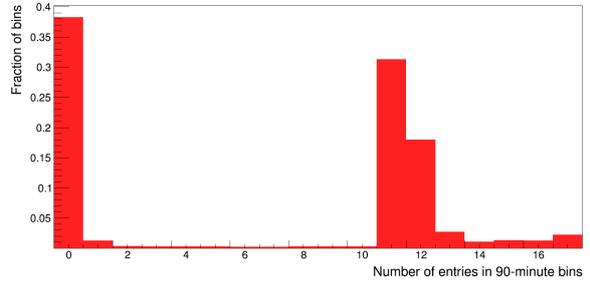
(b) $\delta t = 45$ min



(c) $\delta t = 60$ min



(d) $\delta t = 75$ min

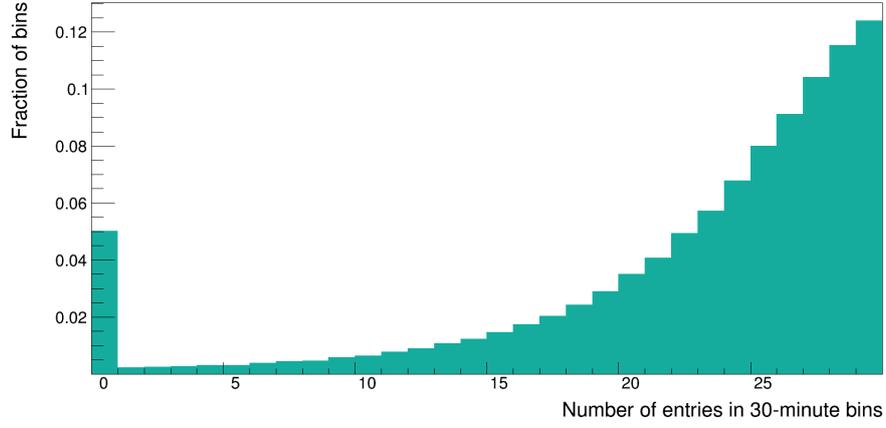


(e) $\delta t = 90$ min

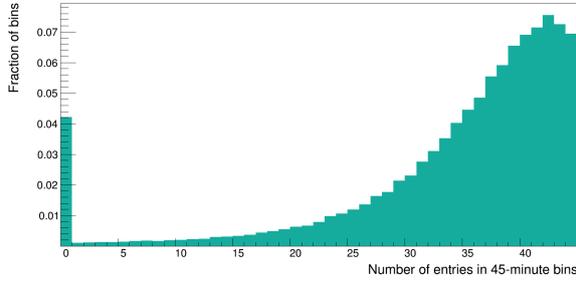
Figure 6.8: Distribution of Scaler data entries per averaging time-window, varying from $\delta t = 30$ min to $\delta t = 90$ min.

the Scaler data are calculated and the same resulting portion of data is correspondingly set as the minimal requirement for number of entries in the Solar Wind data. Consider $\delta t = 30$ min as an example: The threshold value of the Scaler data is 3 entries, which provides $3 \cdot 5$ min = 15 min of data, which is 50% of the complete time span. Hence, a minimum of 15 entries is set as a requirement for Solar wind data due to the given 1-minute averaged data. All minimal required entries for both datasets are listed in Table 6.4.

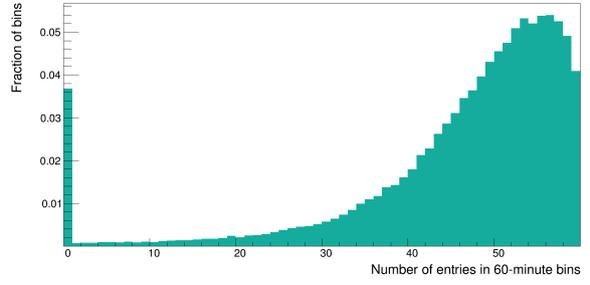
6. Independent analysis of the anticorrelation



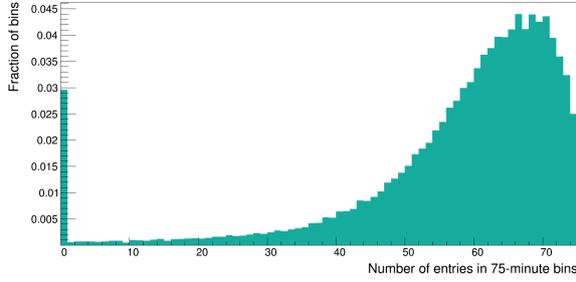
(a) $\delta t = 30$ min



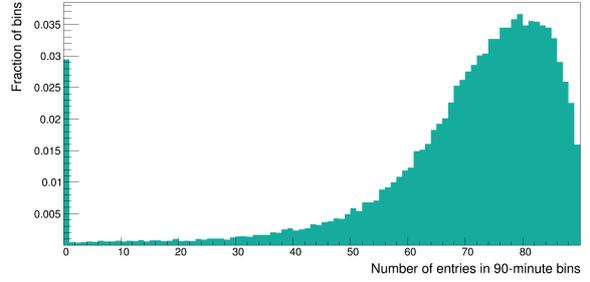
(b) $\delta t = 45$ min



(c) $\delta t = 60$ min



(d) $\delta t = 75$ min



(e) $\delta t = 90$ min

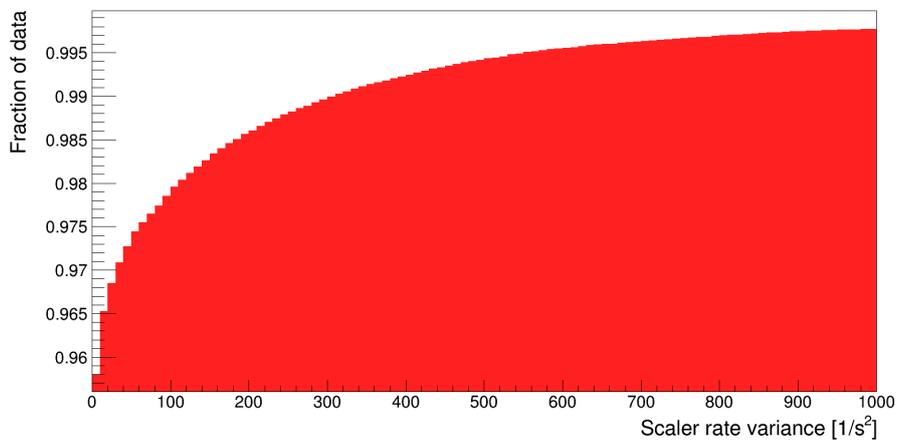
Figure 6.9: Distribution of Solar wind velocity data entries per averaging time-window, varying from $\delta t = 30$ min to $\delta t = 90$ min.

Table 6.4: Minimum required entries depending on δt .

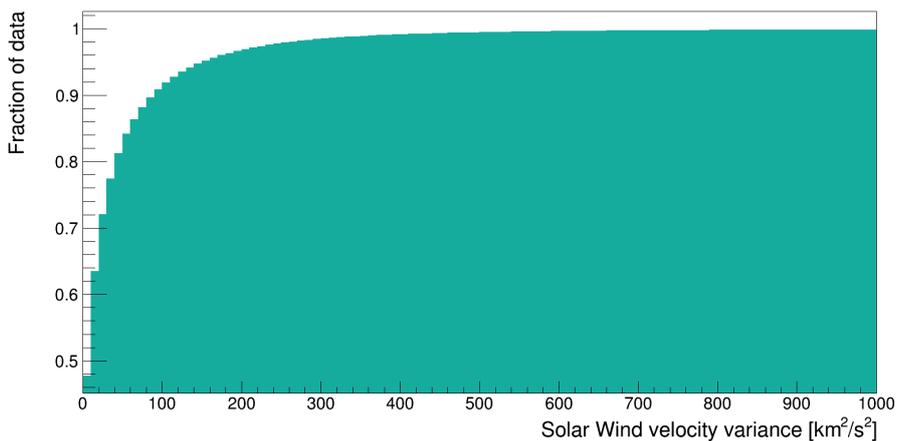
dataset	δt [min]				
	30	45	60	75	90
Scaler	3	4	7	9	11
Solar wind velocity	15	20	35	45	55

6.4. Variance analysis

After all above considerations related to time averaging and data cleaning the distribution of the variance from all data periods is studied to see if it behaves reasonably and to see if there are any effects left which cause irregularly high noise. In Figure 6.10 cumulative distributions of the variances of both datasets are displayed. In general, the variance behaves very smoothly, there are no sudden gains. About 98% of Scaler data has a variance less than $100 \frac{1}{s^2}$ and 90% of Solar Wind velocity data has a variance less than $100 \frac{km^2}{s^2}$. The plots do not show all entries; with decreasing δt the number of entries in overflow increases from 3 up to 200 for the Scaler data, and ranges between 74 and 97 for the Solar Wind velocity data.



(a)

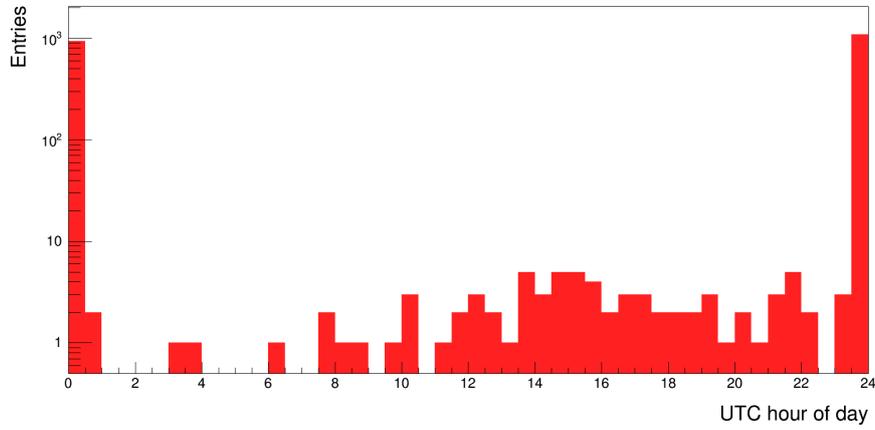


(b)

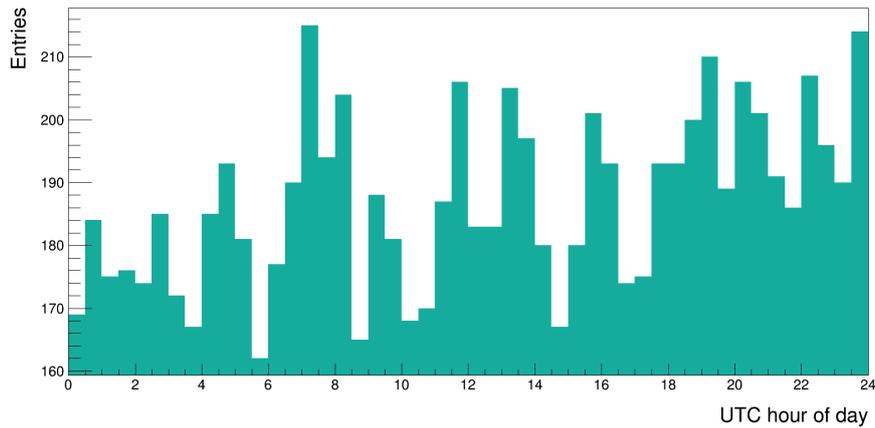
Figure 6.10: Cumulative distributions of the variance of (a) the Scaler data and (b) the Solar Wind velocity data.

6. Independent analysis of the anticorrelation

However, the big dependence of the Scaler data on the hour of day, as discussed in Section 6.2.2, gives reason to study the daily variation of its variance. The distribution of 30 min data bins with variances higher than $100 \frac{1}{s^2}$ (and $100 \frac{km^2}{s^2}$ respectively) are depicted in Figure 6.11. In Figure 6.11(a) it is clearly visible that the vast majority of high-variance Scaler data is in the hour surrounding midnight. This pattern holds for all values of the averaging length δt . The reason for this behaviour is unclear, but the data is made unnecessarily worse by this effect. On the other hand, the high-variance Solar Wind velocity data shows no clear correlation with the hour of day (see Figure 6.11(b)). Data for other choices of δt can be found in Appendix A.3. Therefore, as a final cut to the data, the first and last average values for each day are removed from the data, as they have vastly more entries with high variances that cannot be explained or corrected.



(a)



(b)

Figure 6.11: Variance distribution depending on time of day for (a) the Scaler data (using a logarithmic scale) and (b) the Solar Wind velocity data ($\delta t = 30$ min).

6.5. Results

With the full data cleaning and selection of Sections 6.1 to 6.4 applied, the correlation analysis can be performed. First, the observables have to be defined:

- $S_{\delta t}$: The Scaler rate, in units of $\frac{1}{s}$, from dataset B after all cuts and averaged over the array and a time span δt (given in minutes).
- $\mu_{S,\delta t}$: The 10-year mean of all corrected Scaler rates $S_{\delta t}$ for a given δt .
- $\delta S_{\delta t} = \frac{S_{\delta t} - \mu_{S,\delta t}}{\mu_{S,\delta t}}$: The fractional deviation of a certain $S_{\delta t,i}$ from $\mu_{S,\delta t}$.
- $v_{SW,\delta t}$: Solar Wind velocity, in units of $\frac{km}{s}$, averaged over a time span δt .

The 27 d solar rotation is an unwanted effect in the data. To get rid of it, again a centered moving average with length of exactly 27 d is computed for every datapoint for both datasets, i.e. every 30 min to 90 min depending on the data averaging length δt . Analogous to the above quantities, the 27 d running mean data $R(S_{\delta t})$, $R(\delta S_{\delta t})$ and $R(v_{SW,\delta t})$ are defined. This leads to two definitions:

- $\delta S_{\delta t}^* = \delta S_{\delta t} - R(\delta S_{\delta t}) = \frac{S_{\delta t} - R(S_{\delta t})}{\mu_{S,\delta t}}$: The difference of the fractional deviations of the Scaler rate and its corresponding centered moving average for a given δt .
- $v_{SW,\delta t}^* = v_{SW,\delta t} - R(v_{SW,\delta t})$: The difference of the Solar Wind velocity and its corresponding centered moving average for a given δt .

The uncertainties of these two quantities can be computed via error propagation. For any running mean R with n_R contributing datapoints x_i , each with an uncertainty Δx_i themselves, the uncertainty of R , ΔR , is given by Equation (6.1). The uncertainty of $\mu_{S,\delta t}$, $\Delta \mu_{S,\delta t}$, with n_μ contributing datapoints and sample standard deviation σ_μ , can be calculated via Equation (6.2). With both $\Delta S_{\delta t}$ and $\Delta v_{SW,\delta t}$ in hand the uncertainties $\Delta(\delta S_{\delta t}^*)$ and $\Delta(v_{SW,\delta t}^*)$ can be calculated as shown in Equations (6.3) and (6.4):

$$\Delta R = \sqrt{\frac{\sum_i \Delta x_i^2}{n_R}} \quad (6.1)$$

$$\Delta \mu_{S,\delta t} = \frac{\sigma_\mu}{\sqrt{n_\mu}} \quad (6.2)$$

$$\Delta(\delta S_{\delta t}^*) = \sqrt{\left(\frac{\Delta S_{\delta t}}{\mu_{S,\delta t}}\right)^2 + \left(\frac{\Delta R(S_{\delta t})}{\mu_{S,\delta t}}\right)^2 + \left(\frac{[S_{\delta t} - R(S_{\delta t})] \cdot \Delta \mu_{S,\delta t}}{\mu_{S,\delta t}^2}\right)^2} \quad (6.3)$$

$$\Delta(v_{SW,\delta t}^*) = \sqrt{\Delta v_{SW,\delta t}^2 + \Delta R(v_{SW,\delta t})^2} \quad (6.4)$$

In Figure 6.12 the time evolution of $\delta S_{\delta t}$ and $v_{SW,\delta t}$ and their corresponding 27 d centered moving average data are depicted for $\delta t = 30$ min (data for other choices of δt can be found in Appendix A.4).

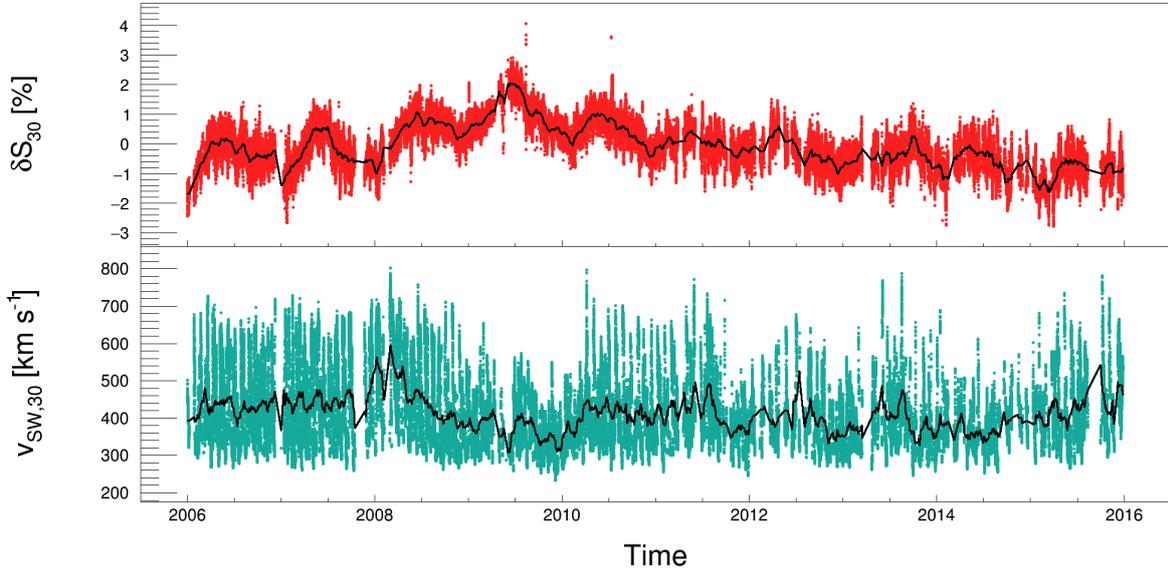


Figure 6.12: Variation in Scaler rate (red points in upper graph) and Solar Wind velocity (blue points in lower graph) including the respective centered moving average of the data (solid black lines) for $\delta t = 30$ min.

One can see that the variation of the Scaler rate is within $\pm 3\%$ of the overall mean rate while the Solar Wind velocity varies on a scale of up to $\pm 300 \frac{\text{km}}{\text{s}}$. The running mean graphs form a good average for the fluctuations of both datasets. To show that a 27 d centered moving average mitigates the contributions from solar rotation, Figure 6.13 shows the dependency of Solar Wind velocity data depending on the day of the solar rotation cycle. To produce these histograms, the full 10 years of data has been divided into 27 d segments and the 30-minute averaged Solar Wind velocity data and their corresponding centered moving average data have been sorted into daily bins by day of the solar rotation cycle. The mean deviation from the respective 27 d mean is calculated for each of these sets of binned data. The upper histogram of Figure 6.13 shows a pattern of three oscillations with magnitudes of up to 15 % of the 27 d mean. The centered moving average data (lower histogram of Figure 6.13) smooths this pattern and shows a maximum deviation of only 1 % from the mean.

6. Independent analysis of the anticorrelation

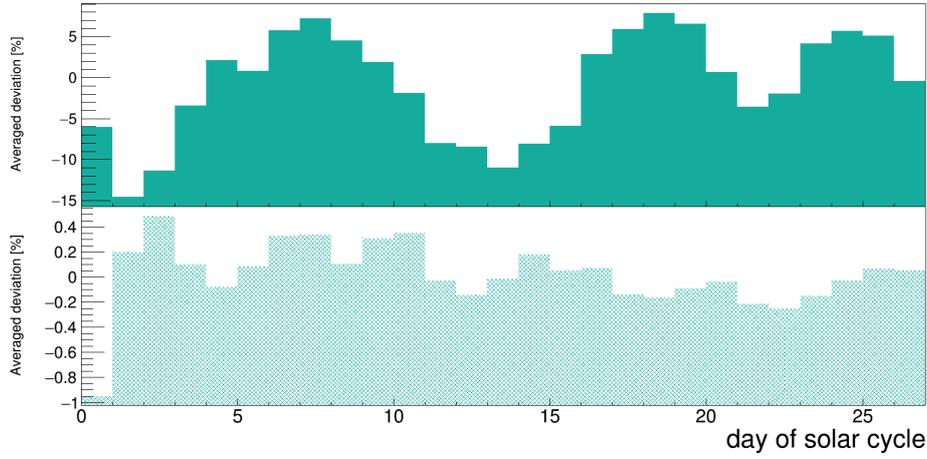


Figure 6.13: 30-minute averaged Solar wind velocity data (upper histogram) and the corresponding 27 d centered moving average of the data (lower histogram) binned according to the day of the solar rotation cycle.

The time evolution of δS_{30}^* and $v_{SW,30}^*$ after subtracting out the centered moving average is shown in Figure 6.14 (data for other choices of δt can be found in Appendix A.4). Both of these datasets show statistical fluctuations around 0. These data will be used to test the anticorrelation.

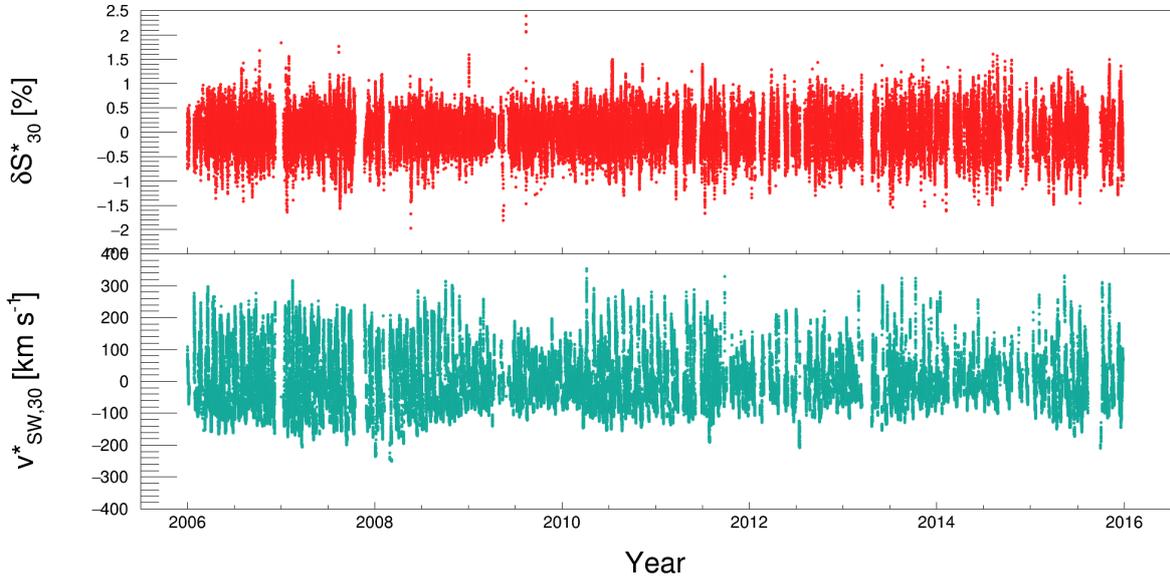


Figure 6.14: Variation in Scaler rate (red points in upper graph) and Solar Wind velocity (blue points in lower graph) after subtracting the respective centered moving average data for $\delta t = 30$ min.

6. Independent analysis of the anticorrelation

In Figure 6.15 a 2D histogram of δS_{30}^* and $v_{\text{SW},30}^*$ is presented (data for other choices of δt can be found in Appendix A.4). Also shown is a profile line for each bin of $v_{\text{SW},30}^*$ representing the mean value and its uncertainty for all δS_{30}^* in that bin. Although a linear fit has not been performed, the trend of the profile line hints at an anticorrelation between both quantities, as it decreases as velocity increases from $-150 \frac{\text{km}}{\text{s}}$ to $250 \frac{\text{km}}{\text{s}}$. Outside of that interval, the statistics are very low and large random fluctuations can be seen in the profiles.

The anticorrelation between δS_{30}^* and $v_{\text{SW},30}^*$ is fit via Equation (6.5). The data overlaid with the result can be seen in Figure 6.16 below, while the fit results for the range of δt values can be seen in Table 6.5 (data for other choices of δt can be found in Appendix A.4).

$$f(v_{\text{SW},\delta t}^*) = m_{\delta t} \cdot v_{\text{SW},\delta t}^* + b_{\delta t} \quad (6.5)$$

Table 6.5: Linear fit parameter results depending on δt .

δt [min]	$m_{\delta t} [\frac{10^{-4} \%}{\text{km s}^{-1}}]$	$b_{\delta t} [10^{-3} \%]$	χ^2/n_{df}
30	-7.99 ± 0.01	3.9 ± 0.1	104.02
45	-8.72 ± 0.02	3.4 ± 0.2	86.77
60	-8.03 ± 0.02	4.6 ± 0.2	115.02
75	-8.31 ± 0.02	4.0 ± 0.2	106.53
90	-8.52 ± 0.03	6.1 ± 0.2	108.40

The calculated uncertainties for most datapoints are very small and the linear fits return very large reduced χ^2 values. However, the slopes of the anticorrelation of Scaler rate and Solar Wind velocity have small uncertainties and similar values for each δt which suggests stability of the result. To be able to specify a significance for the anticorrelation, a constant $c_{\delta t}$ is fit to the same data. The results for the range of δt values can be seen in Table 6.6. The constants also have small uncertainties, but their reduced χ^2 values are always higher, and by that, worse, than the corresponding values of the linear fit. This shows that the linear fit is slightly better compatible with the data, but both models (anticorrelation and constant) would fail a hypothesis test due to the reduced χ^2 values above 100.

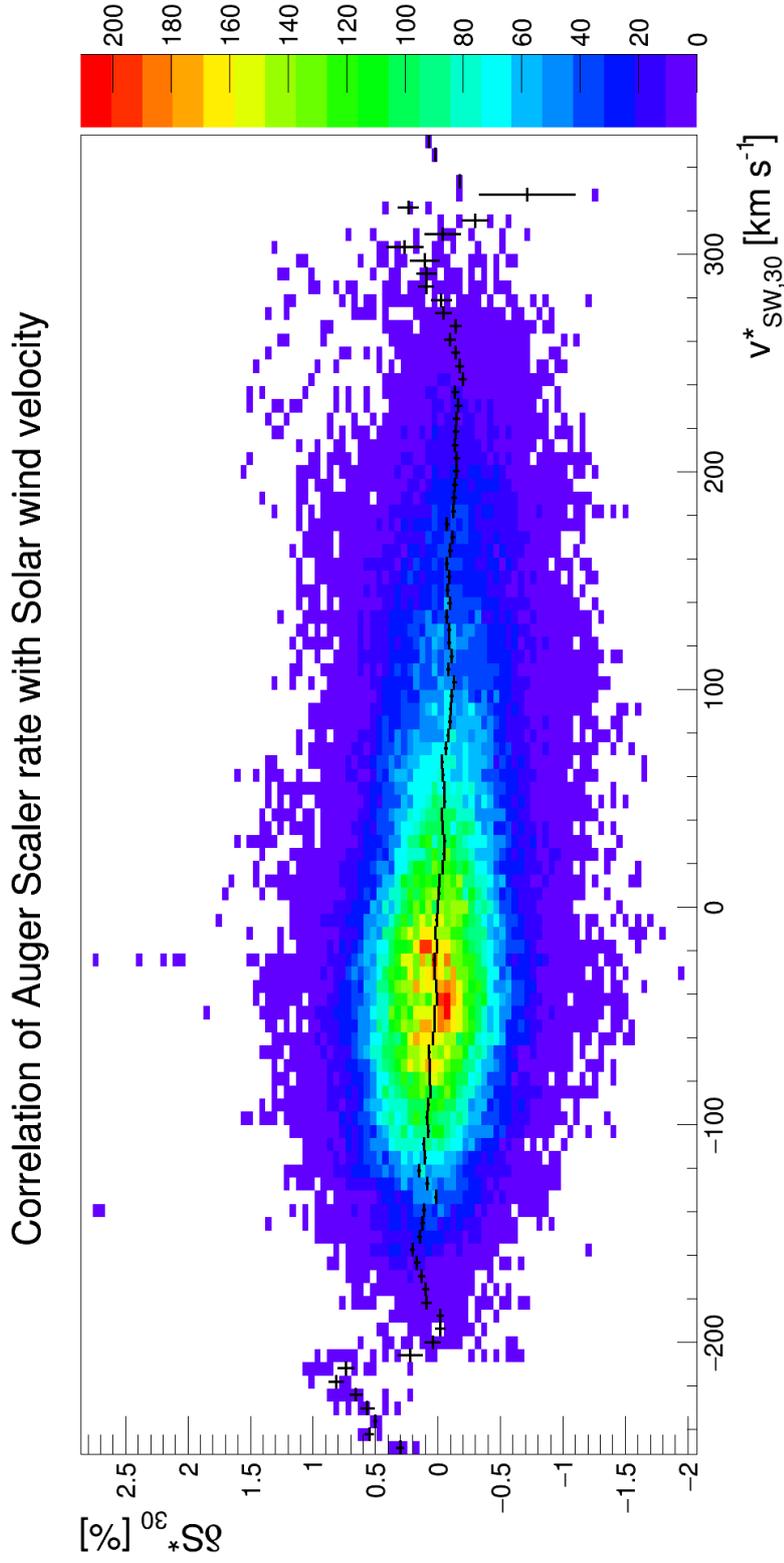


Figure 6.15: 2D histogram of δS_{30}^* and $v_{SW,30}^*$ with profile line (black) for each bin in $v_{SW,30}^*$.

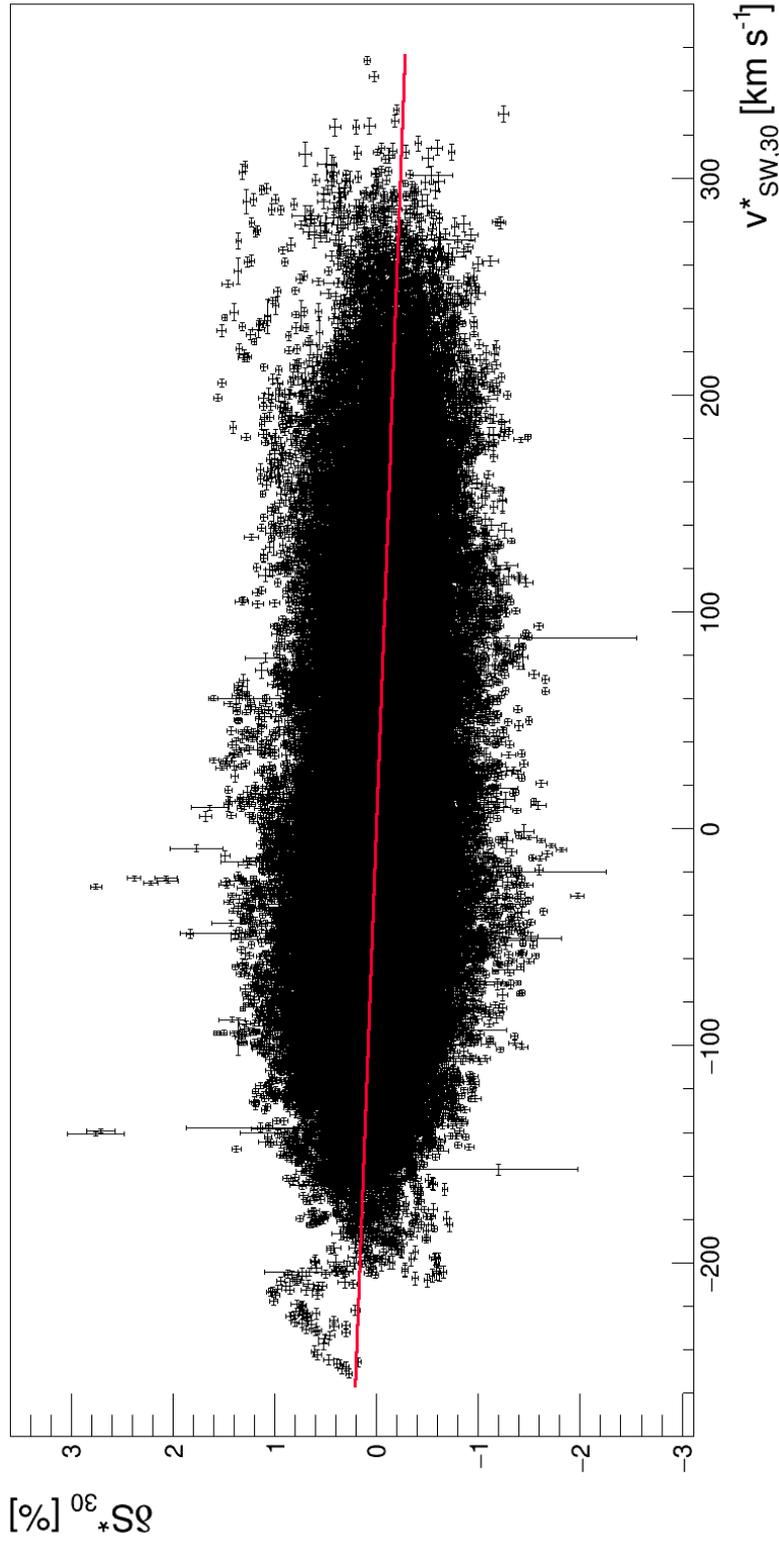


Figure 6.16: Correlation of δS_{30}^* and $v_{SW,30}^*$ with linear fit (red line).

Table 6.6: Constant fit parameter results depending on δt .

δt [min]	$c_{\delta t}[10^{-3} \text{ \%}]$	χ^2/n_{df}
30	4.0 ± 0.1	107.52
45	3.4 ± 0.2	90.00
60	4.9 ± 0.2	119.05
75	4.3 ± 0.2	110.07
90	6.5 ± 0.2	112.58

Finally, the data is sorted by its $v_{\text{SW},dt}^*$ values in ascending order and binned so that the sum of times covered by each bin is exactly 43 days like in the GRAPES-3 analysis. As an example, for $\delta t = 30$ min a total of $\frac{43 \cdot 1440 \text{ min}}{30 \text{ min}} = 2064$ datapoints go in each bin. The data is then fit with both a linear function and a constant. The data for $\delta t = 30$ min is shown in Figure 6.17 (data for other choices of δt can be found in Appendix A.4).

The uncertainties in δS_{30}^* are a bit too high and the uncertainties in $v_{\text{SW},30}^*$ are so small that they are barely visible. Other than that, the linear fit seems to describe the data well which is supported by the statistical fluctuation around 0 of the corresponding residuals. The constant fit does not seem to be a good model, as less than half of the datapoints hit the constant line within their error bar. To quantify these observations the fit parameter results for the two models can be found in Tables 6.7 and 6.8. In addition to the fit parameters and the reduced χ^2 , the p -value is introduced as a parameter to test the goodness of the fit. Given that the null hypothesis (i.e. linear model or constant model) is true, it states the probability of obtaining a result equal to or more extreme than the observed result. In other words, the p -value is the probability of a ‘type I error’ of rejecting a true statistical hypothesis. Usually, models yielding $p < 0.05$ are rejected.

Table 6.7: Linear fit parameter results for binned data depending on δt .

δt [min]	$m_{\delta t}[\frac{10^{-4} \text{ \%}}{\text{km s}^{-1}}]$	$b_{\delta t}[10^{-3} \text{ \%}]$	χ^2/n_{df}	p
30	-6.9 ± 0.7	1 ± 7	0.33	0.999994
45	-7.4 ± 0.8	2 ± 7	0.16	$1 - 7 \cdot 10^{-11}$
60	-7.1 ± 0.7	1 ± 6	0.32	0.999990
75	-7.1 ± 0.8	0.3 ± 6.7	0.24	$1 - 3 \cdot 10^{-7}$
90	-5.7 ± 0.7	4 ± 6	0.95	0.564937

6. Independent analysis of the anticorrelation

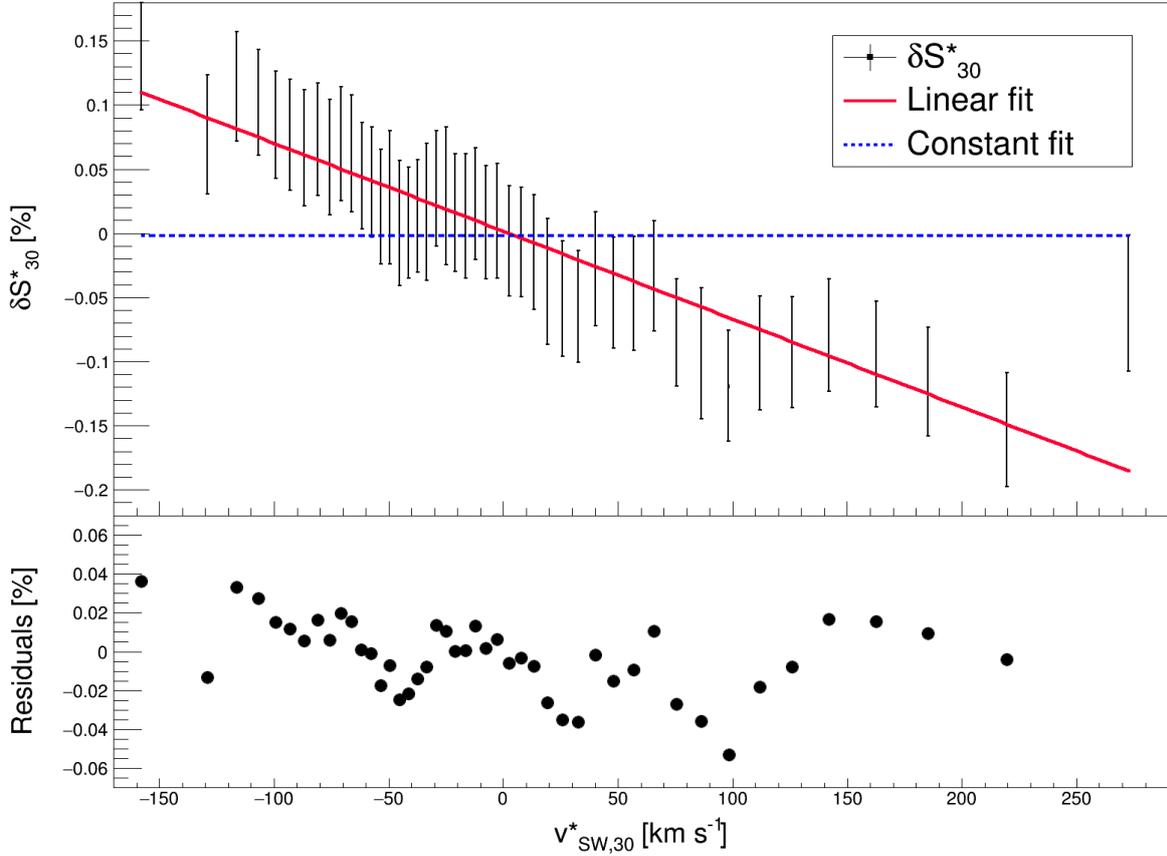


Figure 6.17: Correlation of δS_{30}^* and $v_{\text{SW},30}^*$ with linear fit (red line) and constant fit (blue dotted line). Residuals (black dots) are shown for the linear fit.

Table 6.8: Constant fit parameter results for binned data depending on δt .

δt [min]	$c_{\delta t} [10^{-3} \%]$	χ^2/n_{df}	p
30	-1 ± 7	2.36	$7.18 \cdot 10^{-7}$
45	-2 ± 7	1.84	$5.07 \cdot 10^{-4}$
60	-2 ± 6	3.02	$1.82 \cdot 10^{-10}$
75	-1 ± 6	2.46	$7.85 \cdot 10^{-7}$
90	1 ± 6	2.67	$5.44 \cdot 10^{-8}$

6. Independent analysis of the anticorrelation

First, the slopes $m_{\delta t}$ of the linear fit each have about a 10% uncertainty and they agree with each other within the uncertainty ranges. The reduced χ^2 is a bit below 1 due to the too large uncertainties in δS_{30}^* . The p -values are well above 0.05 so that the anticorrelation model cannot be rejected. In summary, the linear model describes the data well.

Second, the constants of the constant fit model have uncertainties of several 100% and reduced χ^2 values above 2 which shows the instability of the results. The highest p -value of the constant model corresponds to a significance of ~ 3.3 of the standard deviations so that the constant model can be rejected.

Finally, the end result for the slope of the anticorrelation is calculated as weighted mean of the 5 fit parameter results and their uncertainties. The parameter δt of the analysis is interpreted as a systematic uncertainty. The standard deviation for the five different slope values (cf. Table 6.7) is added to the uncertainty of the end result:

$$m_{\text{corr}}^{\text{Scaler,B}} = [-6.8 \pm 0.3 \text{ (stat.)} \pm 0.7 \text{ (syst.)}] \frac{10^{-4}\%}{\text{km}\cdot\text{s}^{-1}} \quad (6.6)$$

Although the anticorrelation found via this independent method is smaller than the slope determined with Scaler dataset A, this result can be well trusted as many systematic uncertainties have been factored in. In conclusion, the anticorrelation of Scaler data and Solar Wind velocity data has been verified.

7. Conclusion and outlook

Due to the effects of the Solar Wind on the propagation of cosmic rays traveling near our solar system, an anticorrelation between the cosmic ray flux on Earth and the Solar Wind velocity is expected. In this analysis, low-energy ‘Scaler’ mode data from the Pierre Auger Observatory is compared to Solar Wind velocity data, provided by NASA OMNIWeb, to test for this predicted anticorrelation.

Two methods and datasets were used in this study. First, the same analysis that was used for the published GRAPES-3 study has been performed on a basic Scaler dataset to directly compare the results of both experiments. In this, periodic effects like the 27-day solar rotation and the 11-year magnetic cycle have been corrected and Forbush decreases have been excluded from the data. The GRAPES-3 data yielded an anticorrelation of $m_{\text{corr}}^{\text{GRAPES}} = [-13.3 \pm 0.7 \text{ (stat.)}] \cdot 10^{-4}\% \text{ s km}^{-1}$. The result for Scaler data is $m_{\text{corr}}^{\text{Scaler,A}} = [-9.0 \pm 0.6 \text{ (stat.)}] \cdot 10^{-4}\% \text{ s km}^{-1}$. The values of the slopes do not agree within error ranges, but a linear anticorrelation is confirmed with comparatively low uncertainties. However, with the Scaler of the Pierre Auger Observatory data spanning 10 years (4 years more than the GRAPES-3 data), a lower uncertainty should be obtainable. This motivated the use of a newer Scaler dataset with advanced systematic cuts to check if a more precise result was possible.

In the independent analysis, the Scaler data was corrected for several effects (e.g. daily oscillations, seasonal patterns, weather effects and detector instabilities). The ‘perfect’ time-window for averaging Scaler and Solar Wind velocity data could not be determined and was therefore kept as a parameter. Future studies on that subject could possibly improve the result through the use of a Fourier analysis to provide a feasible result for the best averaging time window.

The final result obtained for the slope of the linear model of the anticorrelation is $m_{\text{corr}}^{\text{Scaler,B}} = [-6.8 \pm 0.3 \text{ (stat.)} \pm 0.7 \text{ (syst.)}] \cdot 10^{-4}\% \text{ s km}^{-1}$. The anticorrelation found via the independent method is smaller than the slope determined with the basic Scaler dataset. However, this result can be well trusted as several systematic effects have been factored in. The null hypothesis of no anticorrelation can be rejected with a significance larger than 3 standard deviations. In conclusion, the anticorrelation of Scaler data and Solar Wind velocity data has been verified.

Appendix A.

In the following pages the variants of the most important plots will be displayed.

A.1. Determining a time window

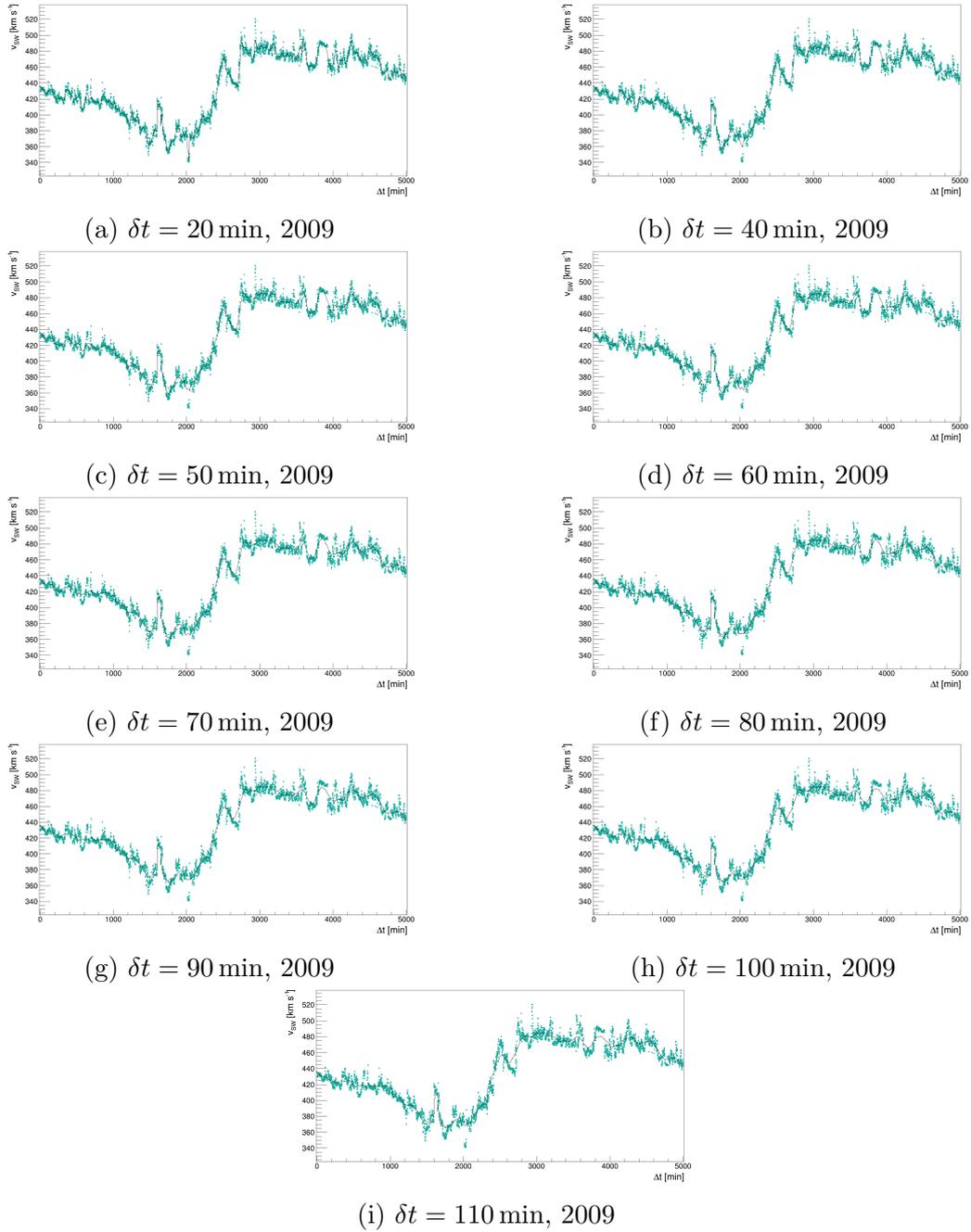


Figure A.1: First 5000 min of a random 27 d sample of Solar Wind velocity data from 2009: 1-minute averages (blue dots) and the running mean of varying length δt (black lines).

Appendix A. Appendix

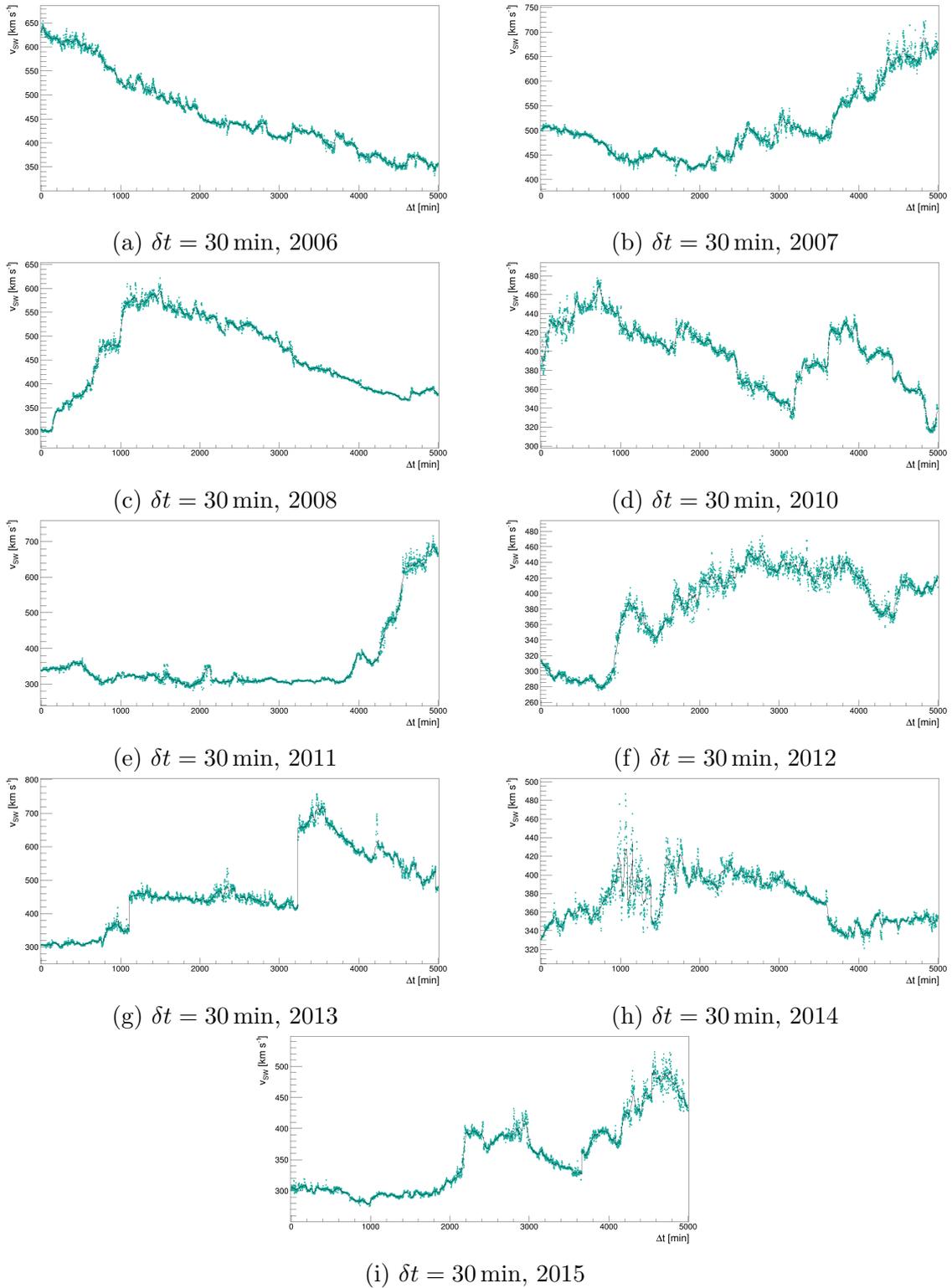
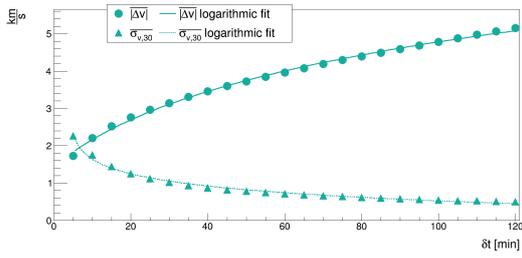
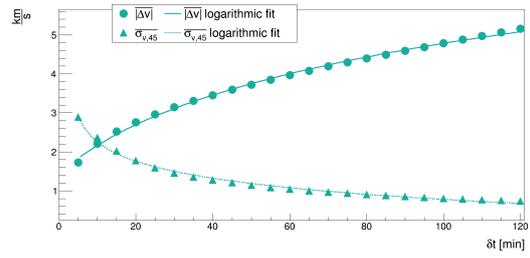


Figure A.2: First 5000 min of random 27 d samples of Solar Wind velocity data from each year from 2006 to 2015: 1-minute averages (blue dots) and the running mean of length $\delta t = 30$ min (black lines).

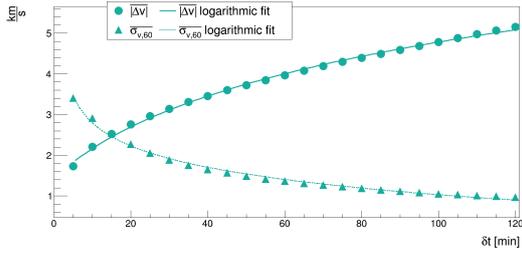
Appendix A. Appendix



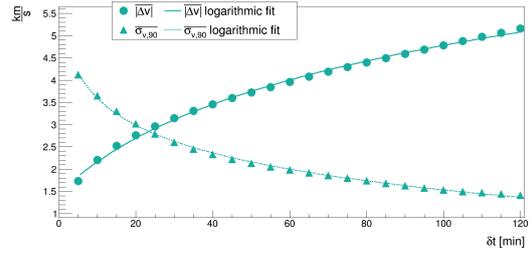
(a) $dt = 30$ min, 2009



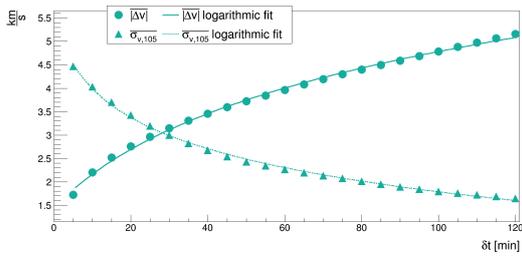
(b) $dt = 45$ min, 2009



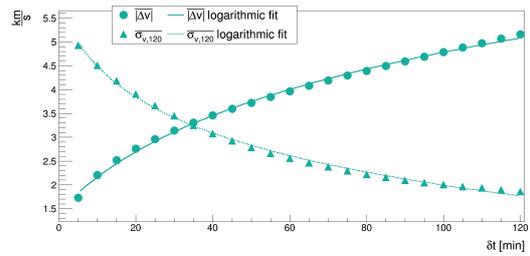
(c) $dt = 60$ min, 2009



(d) $dt = 90$ min, 2009



(e) $dt = 105$ min, 2009



(f) $dt = 120$ min, 2009

Figure A.3: Dependence of $|\overline{\Delta v}|$ (blue dots) and $\overline{\sigma_{v,dt}}$ (blue triangles) on the length of δt for a random Solar Wind velocity sample taken from 2009 with varying length of dt .

Appendix A. Appendix

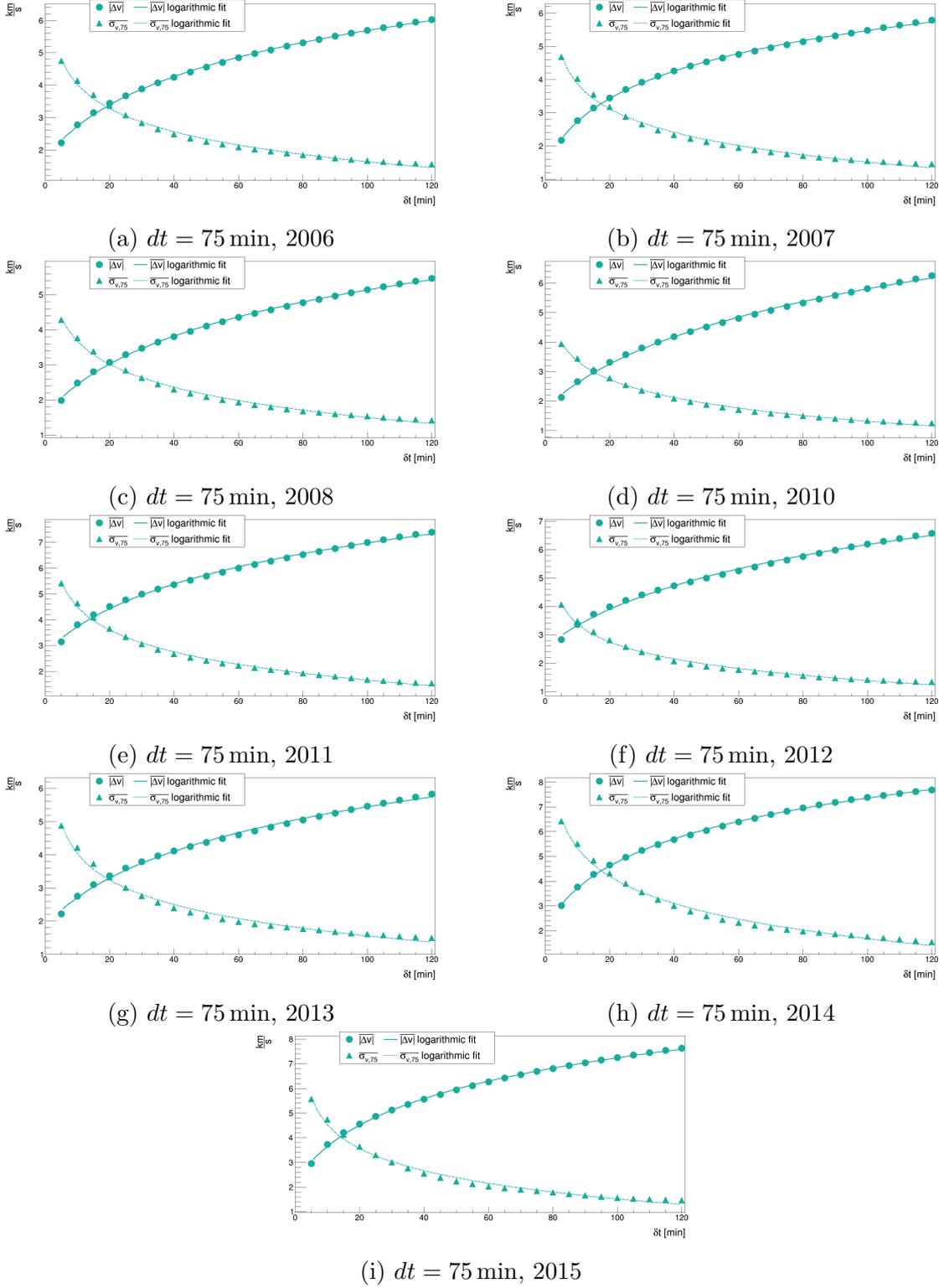


Figure A.4: Dependence of $|\overline{\Delta v}|$ (blue dots) and $\overline{\sigma_{v,75}}$ (blue triangles) on the length of δt for random Solar Wind velocity samples taken from each year from 2006 to 2015 with $dt = 75$ min.

Appendix A. Appendix

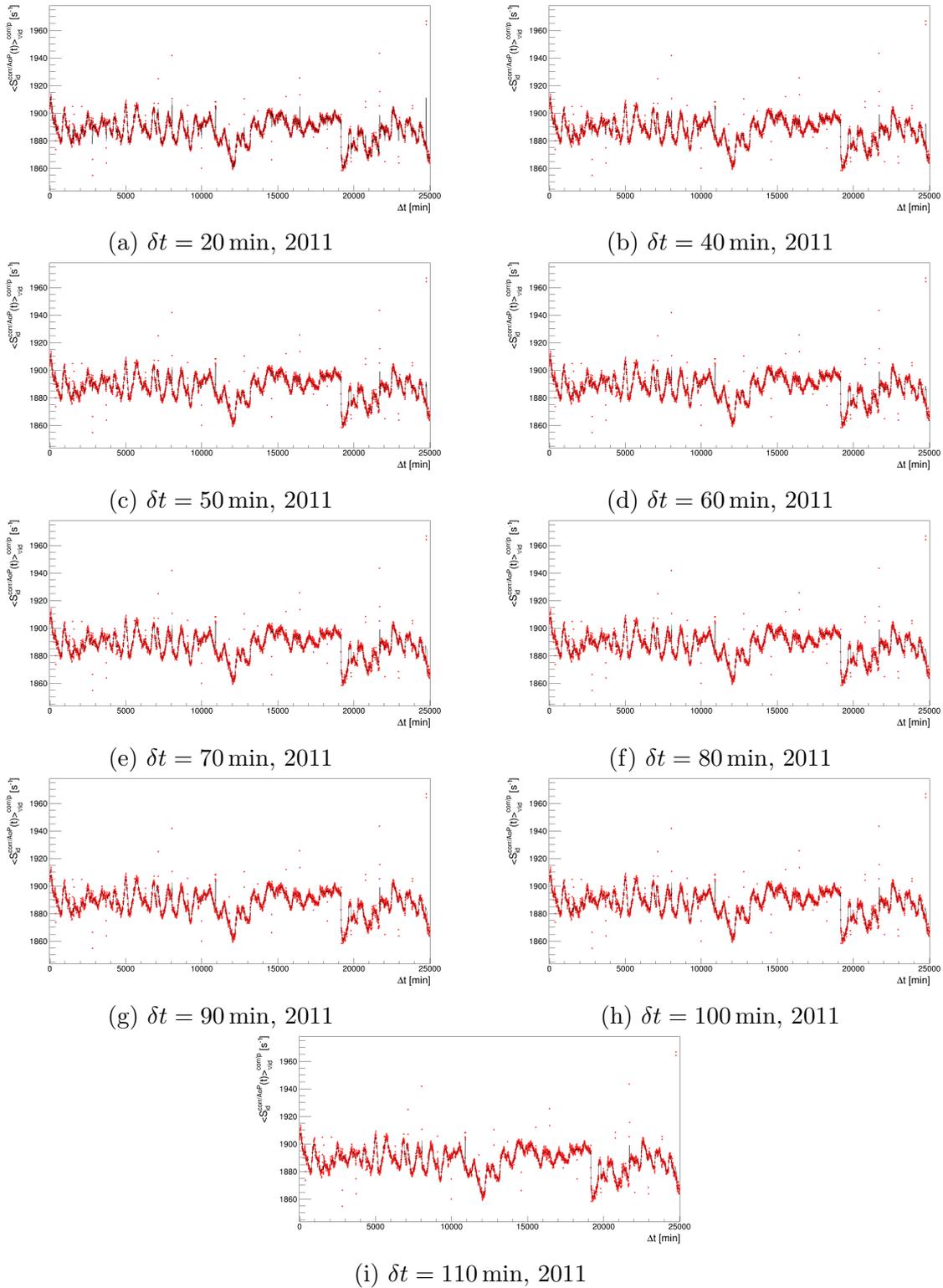


Figure A.5: First 25000 min of a random 27 d sample of Scaler data from 2011: 5-minute averages (red dots) and the running mean of varying length δt (black lines).

Appendix A. Appendix

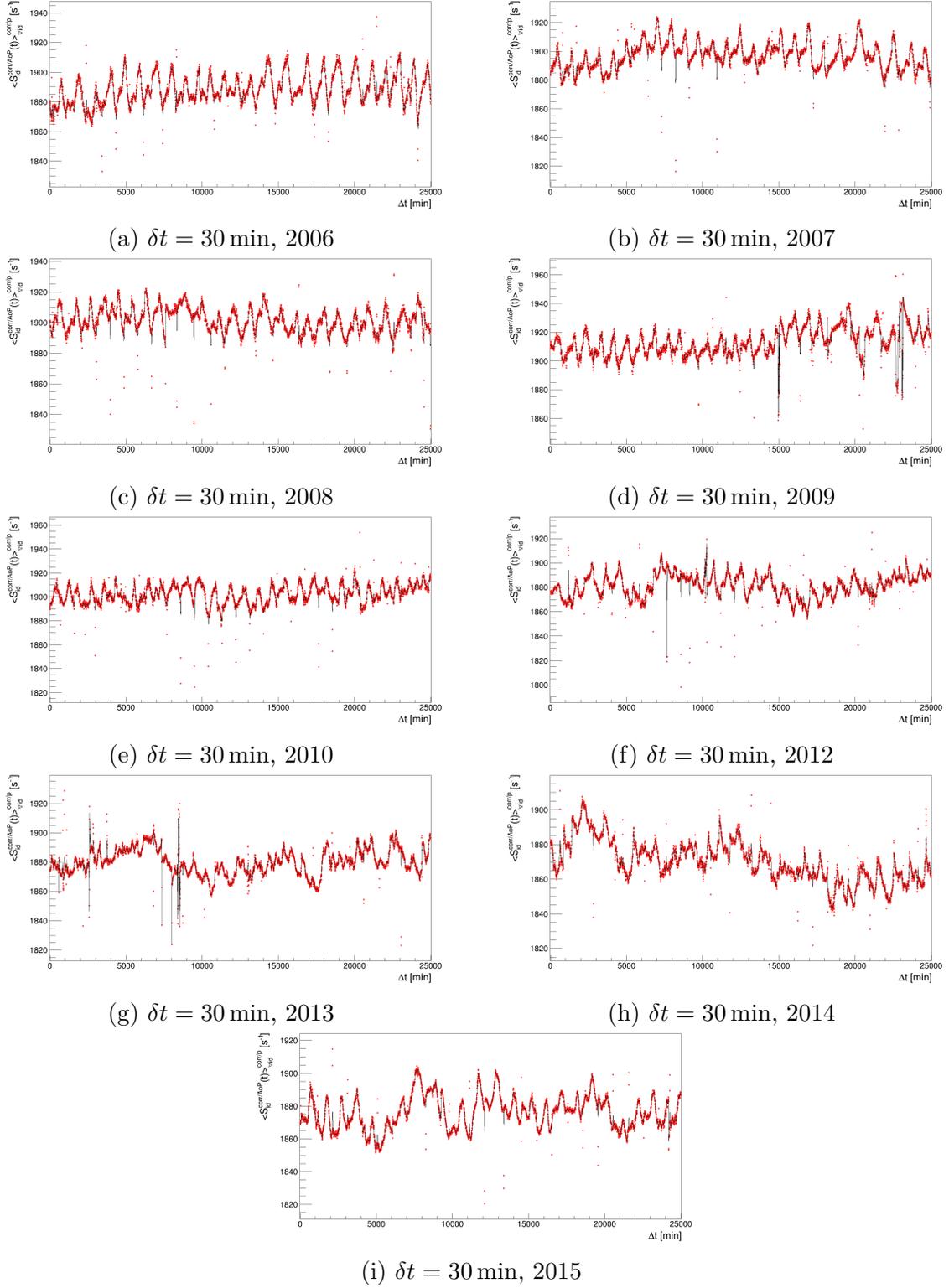


Figure A.6: First 25000 min of random 27 d samples of Scaler data from each year from 2006 to 2015: 5-minute averages (red dots) and the running mean of length $\delta t = 30$ min (black lines).

Appendix A. Appendix

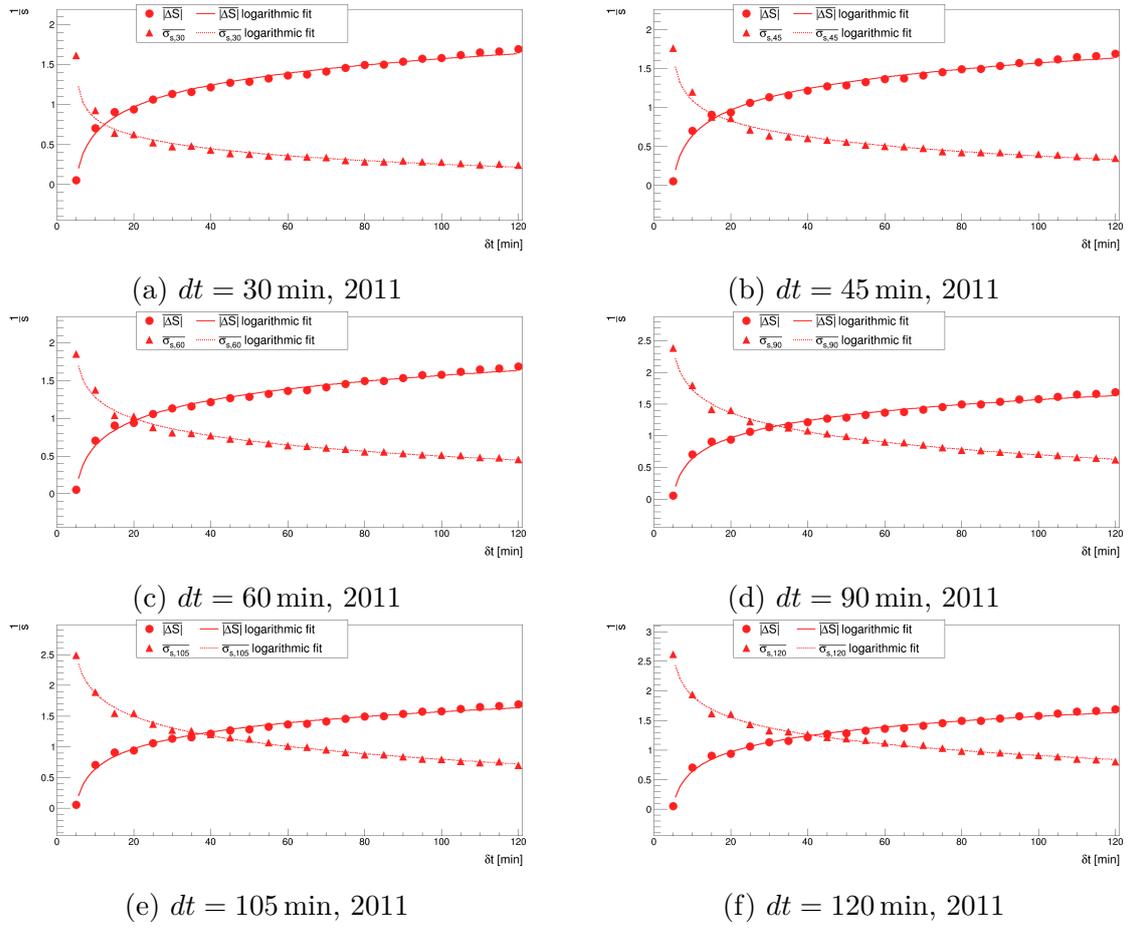


Figure A.7: Dependence of $|\overline{\Delta S}|$ (red dots) and $\overline{\sigma_{S,dt}}$ (red triangles) on the length of δt for a random Scaler data sample taken from 2011 with varying length of dt .

Appendix A. Appendix

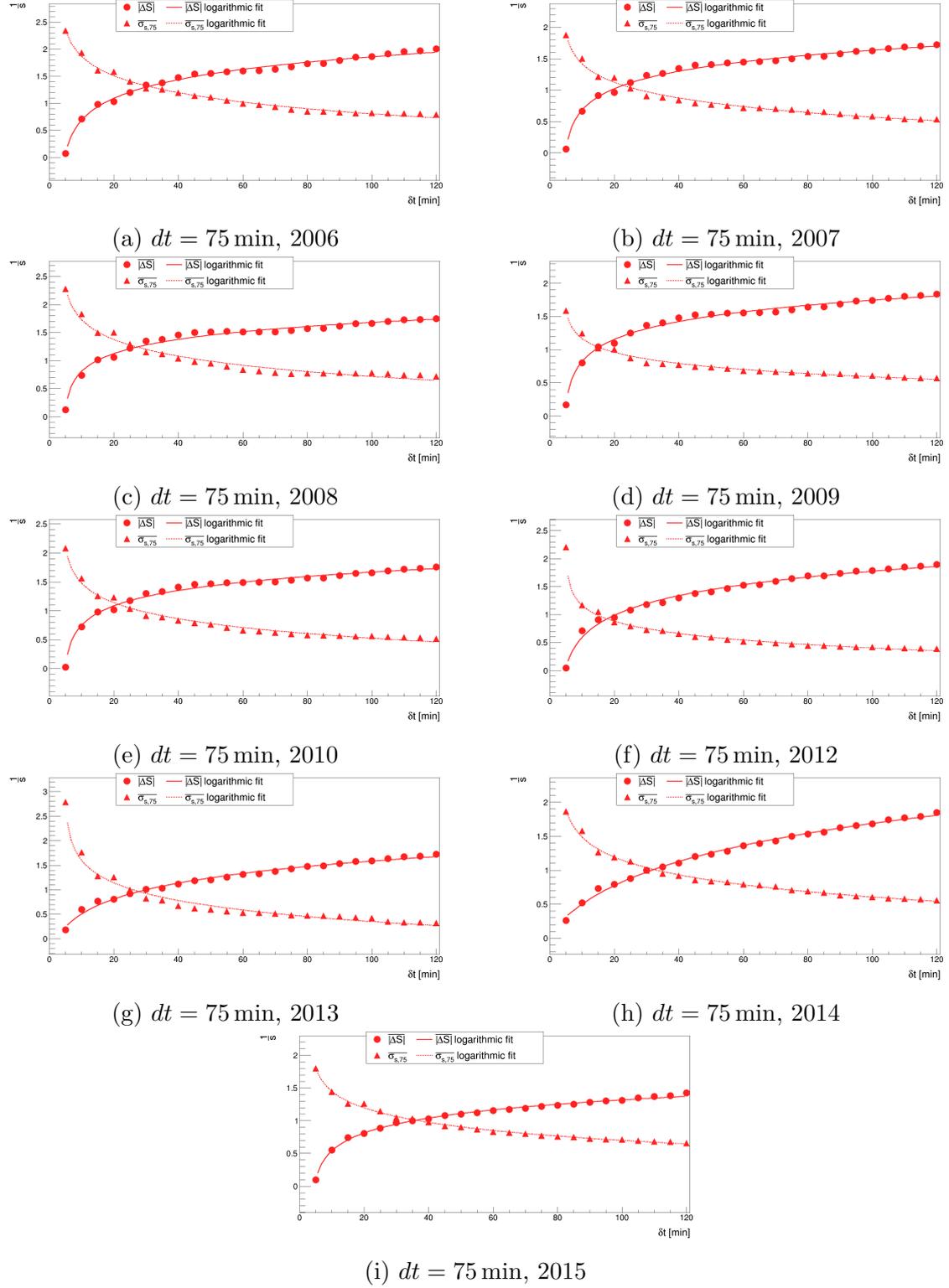


Figure A.8: Dependence of $|\overline{\Delta S}|$ (red dots) and $\overline{\sigma_{S,75}}$ (red triangles) on the length of δt for random Solar Wind velocity samples taken from each year from 2006 to 2015 with $dt = 75$ min.

A.2. Scaler data treatment

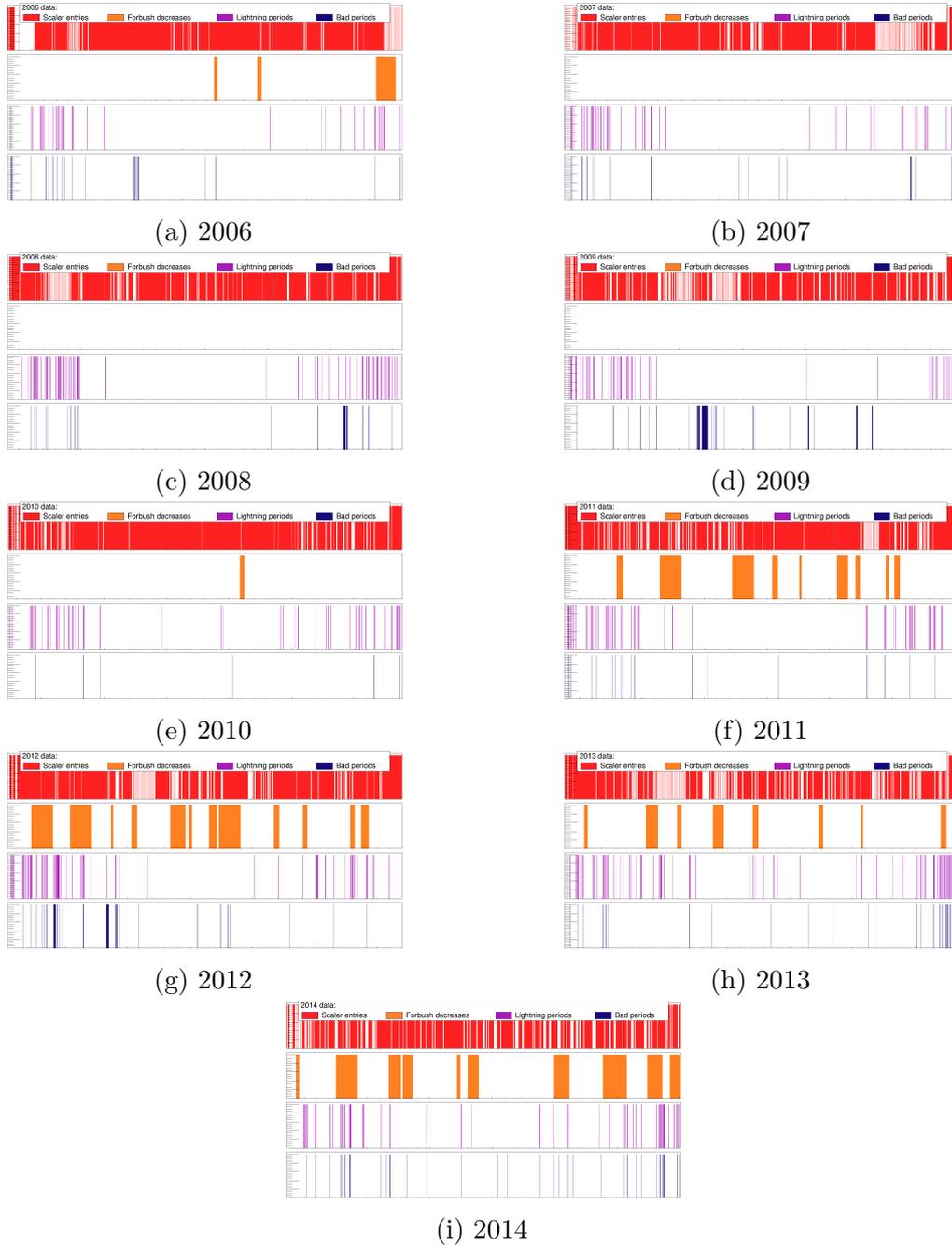


Figure A.9: Scaler data availability (first row, red) and periods removed from the data due to Forbush decreases (second row, orange), lightning periods (third row, purple) and ‘bad periods’ (fourth row, dark blue) from 2006 to 2014.

Appendix A. Appendix

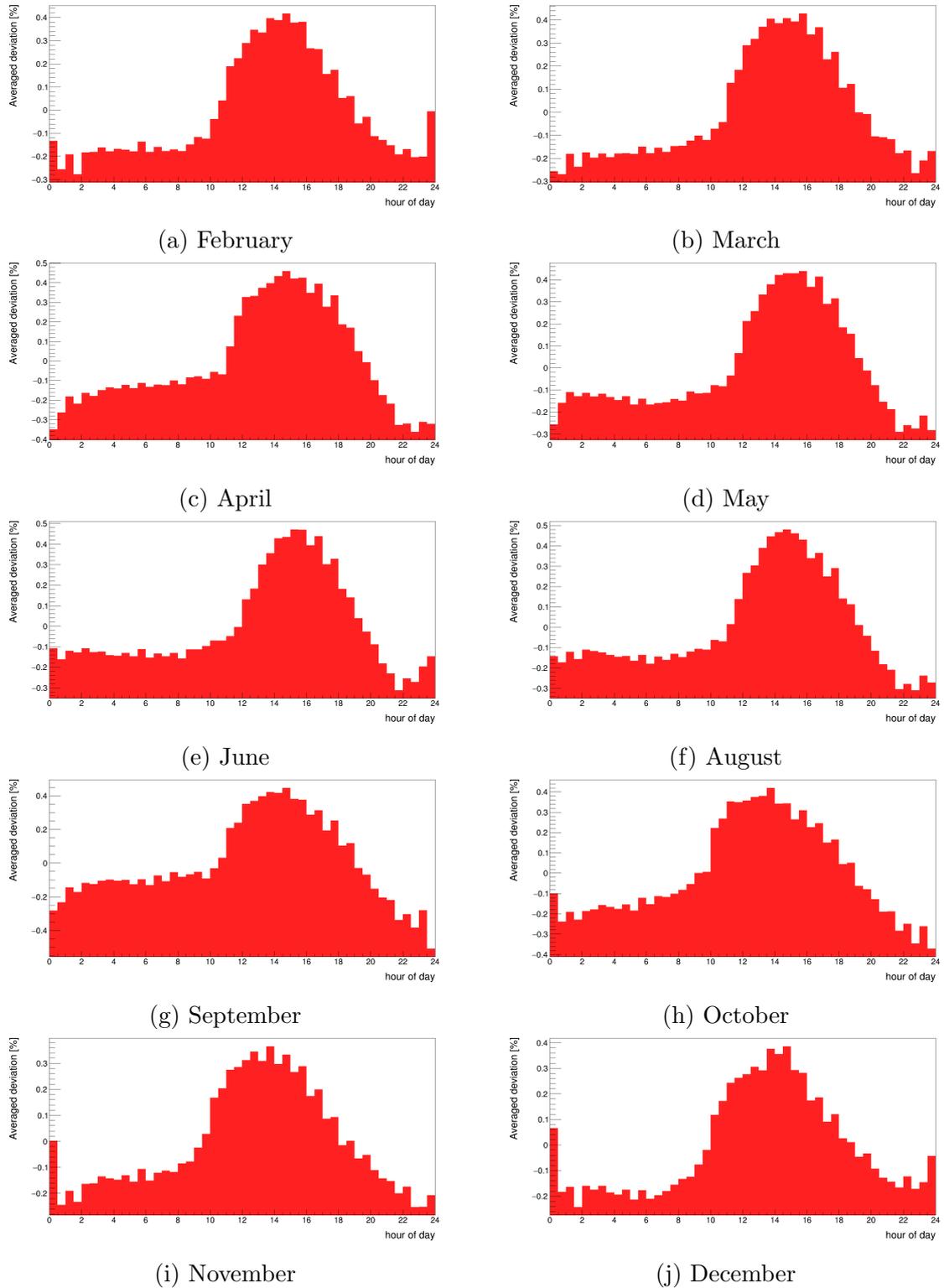


Figure A.10: Average deviations from the daily mean Scaler rates for a chosen time-window of 30 min depending on the month of the year.

A.3. Variance analysis

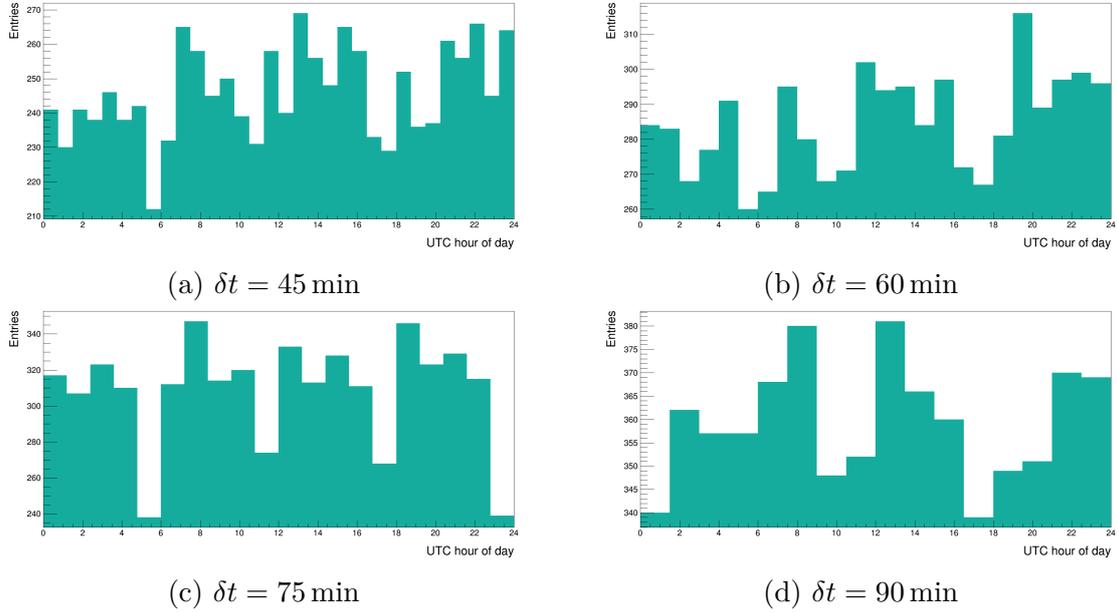


Figure A.11: Variance distribution depending on time of day for the Solar Wind velocity data and varying length of δt .

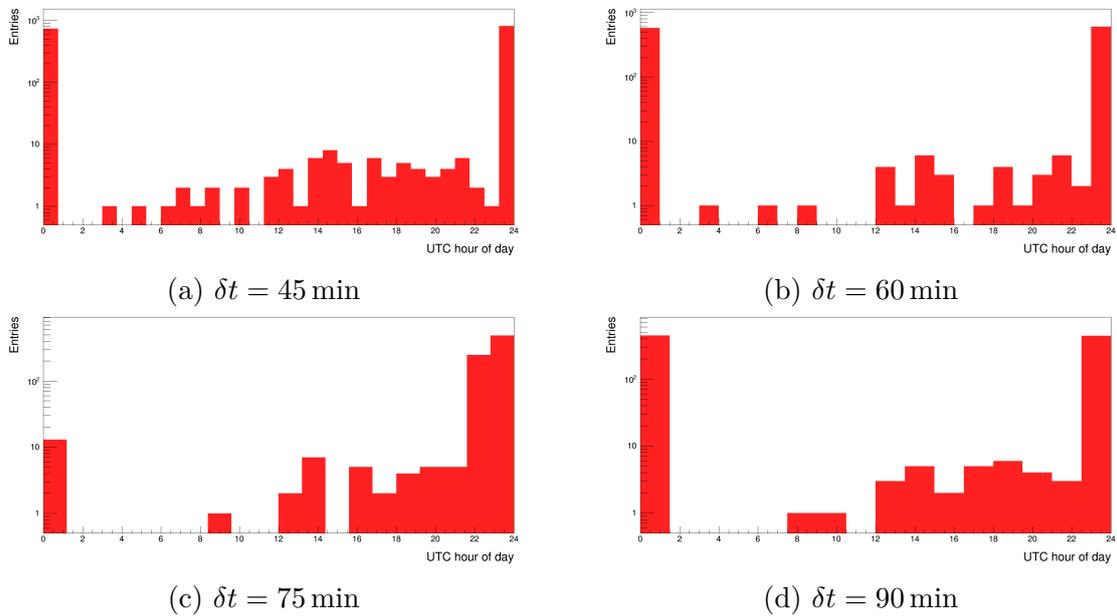
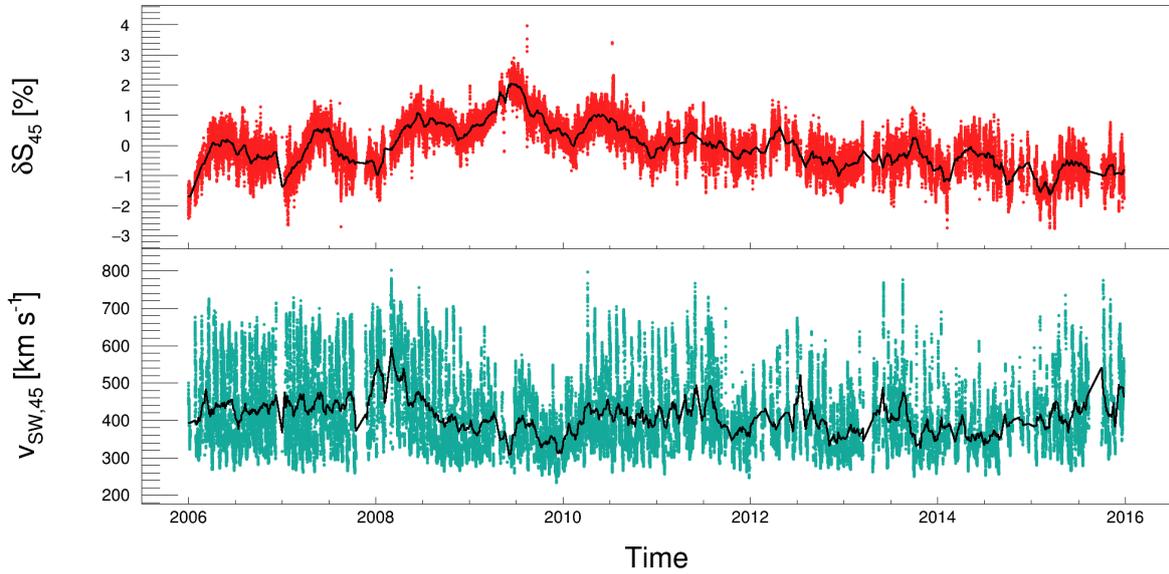
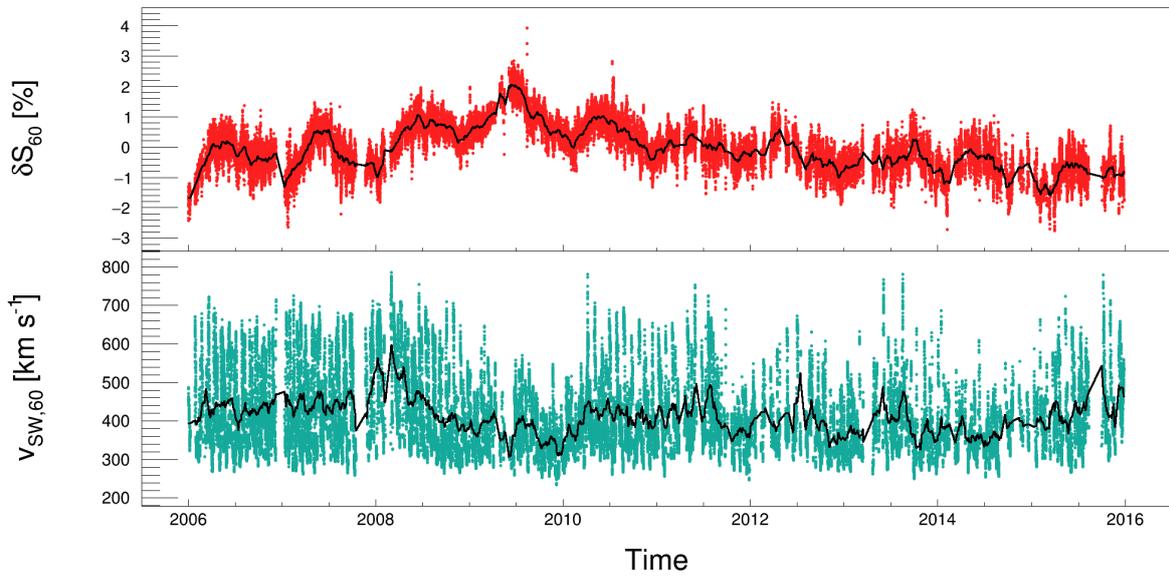


Figure A.12: Variance distribution depending on time of day for the Scaler data and varying length of δt .

A.4. Results

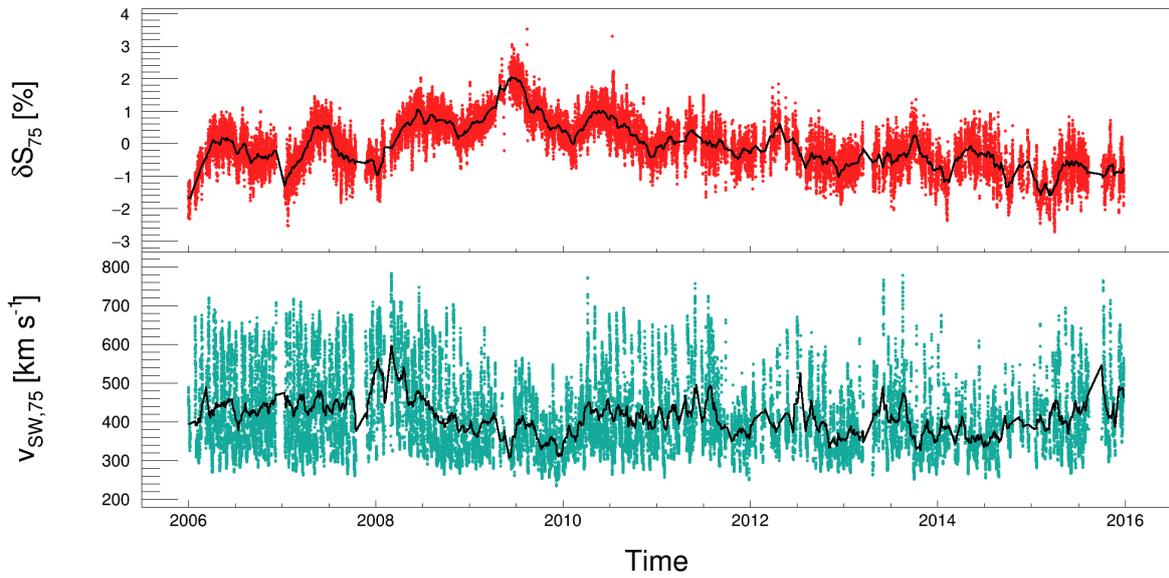


(a) $\delta t = 45$ min

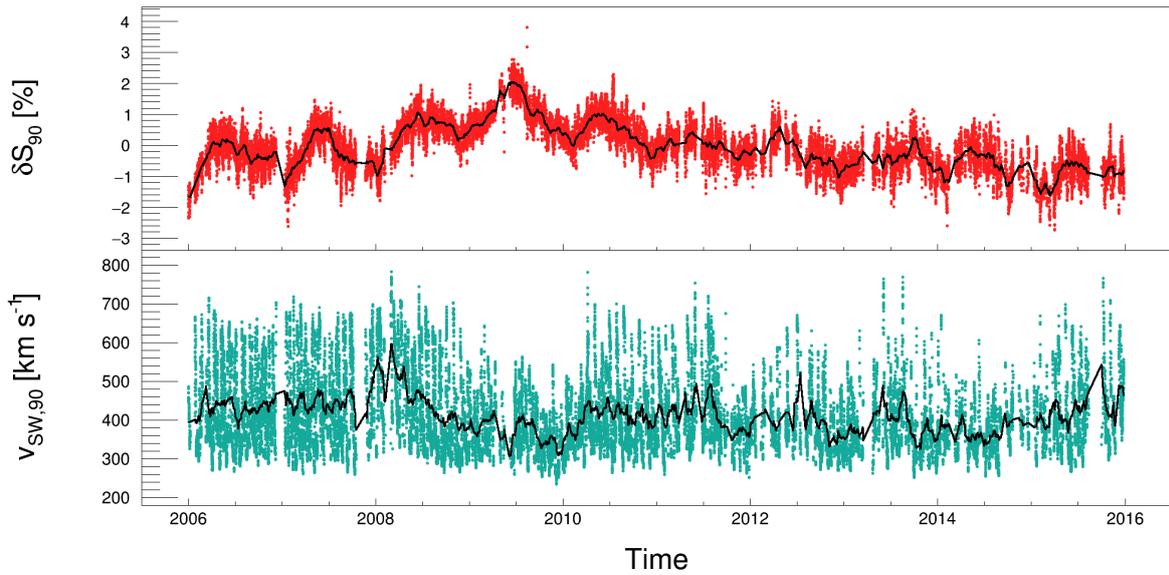


(b) $\delta t = 60$ min

Figure A.13: Variation in Scaler rate (red points in upper pad) and Solar Wind velocity (blue points in lower graph) including the respective centered moving average of the data (solid black lines) for a δt of (a) 45 min and (b) 60 min.

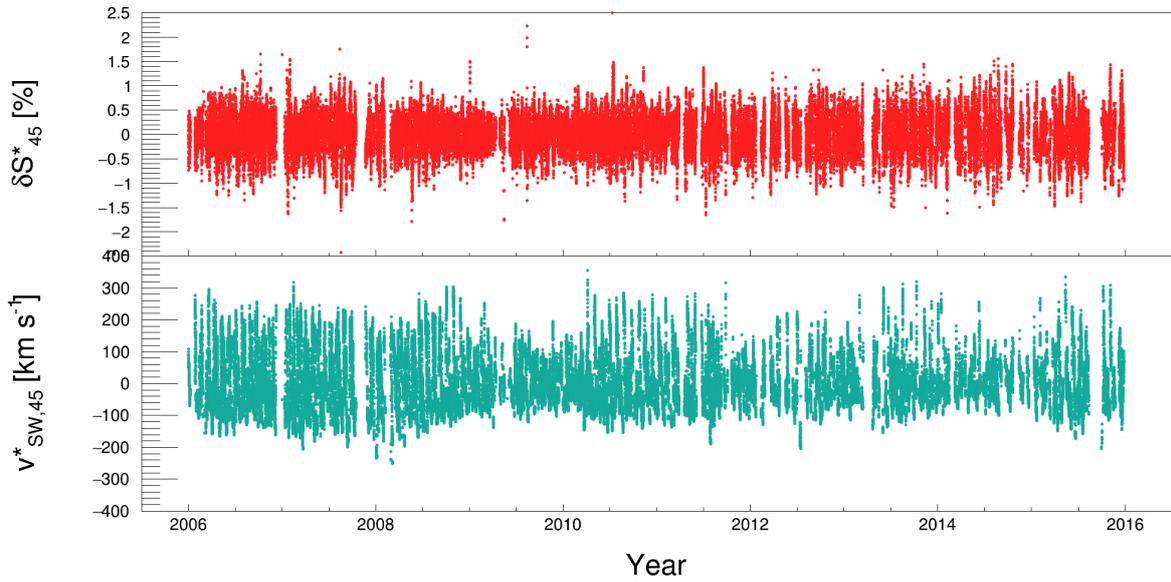


(a) $\delta t = 75$ min

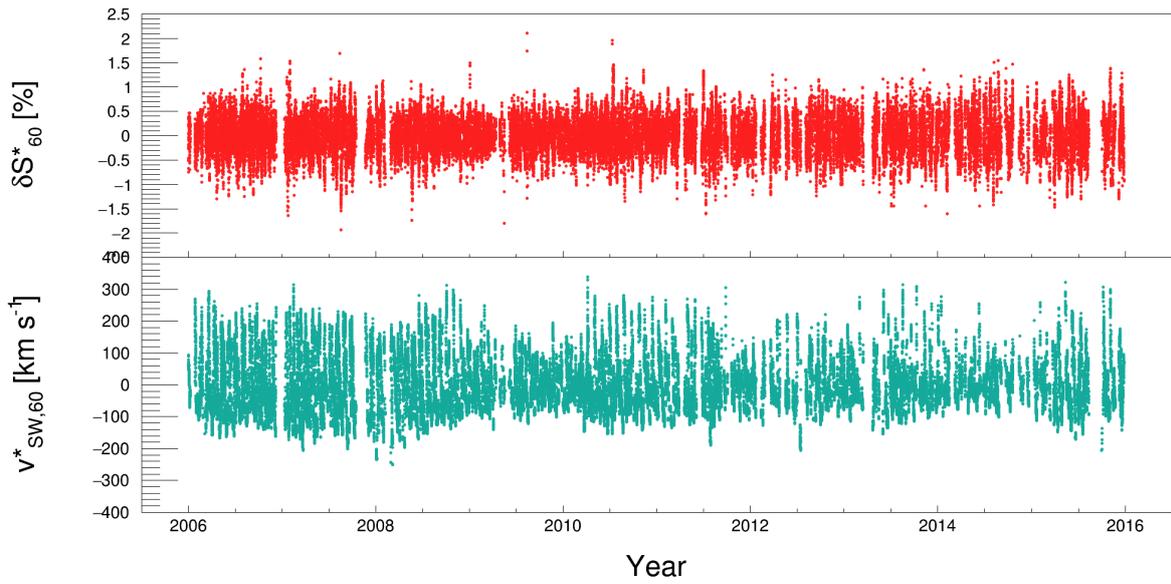


(b) $\delta t = 90$ min

Figure A.14: Variation in Scaler rate (red points in upper pad) and Solar Wind velocity (blue points in lower graph) including the respective centered moving average of the data (solid black lines) for a δt of (a) 75 min and (b) 90 min.

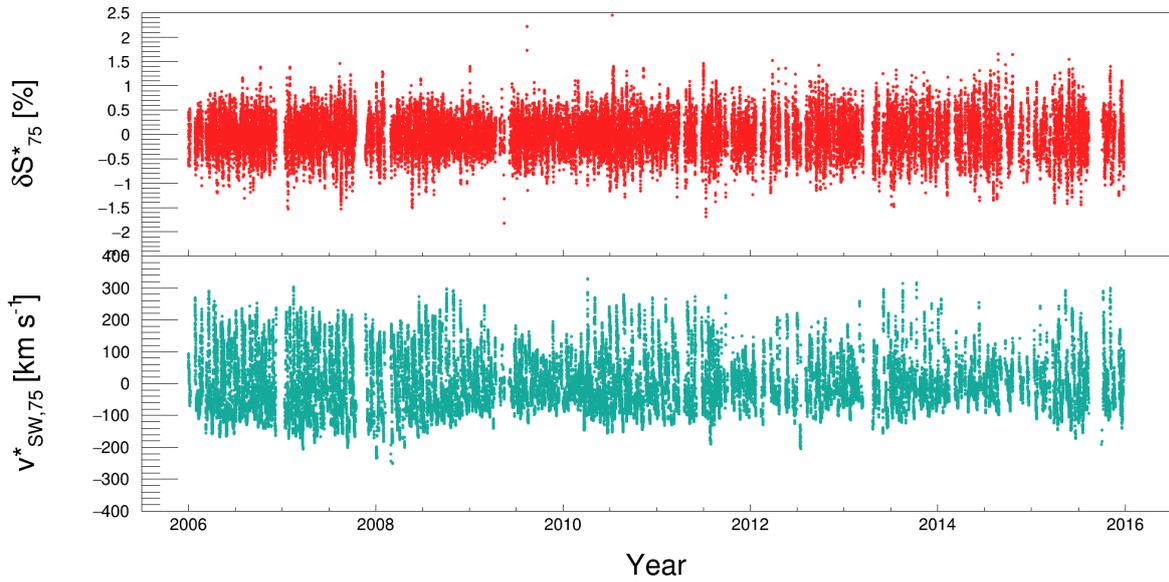


(a) $\delta t = 45$ min

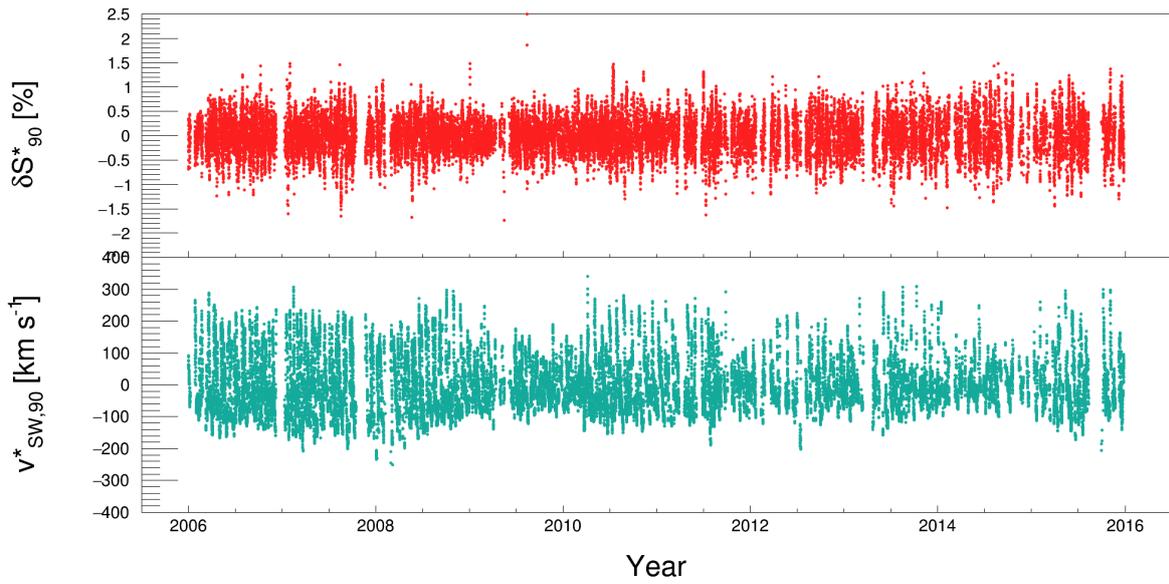


(b) $\delta t = 60$ min

Figure A.15: Variation in Scaler rate (red points in upper graph) and Solar Wind velocity (blue points in lower graph) after subtracting the respective centered moving average data for a δt of (a) 45 min and (b) 60 min.



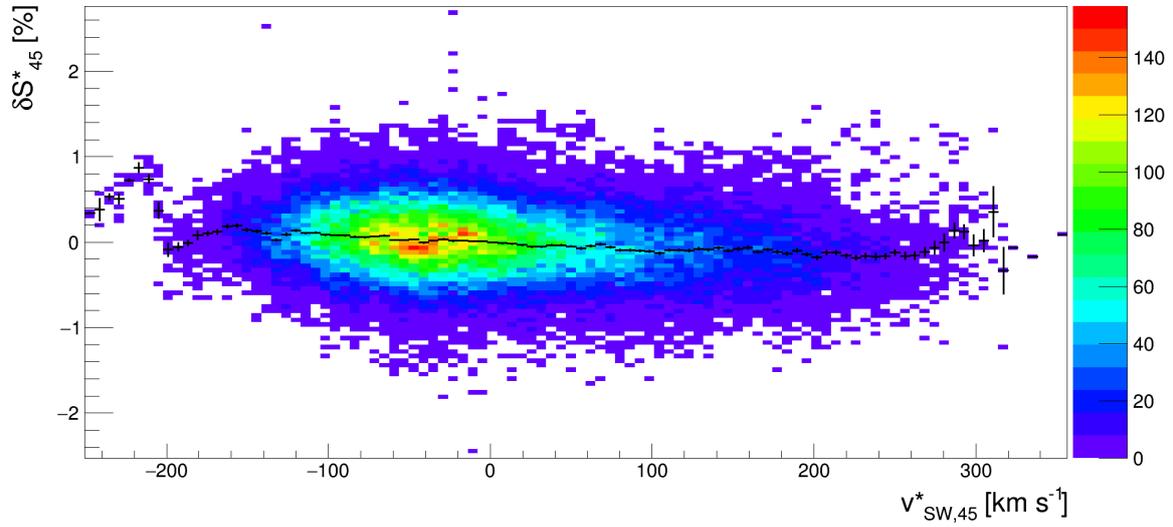
(a) $\delta t = 75$ min



(b) $\delta t = 90$ min

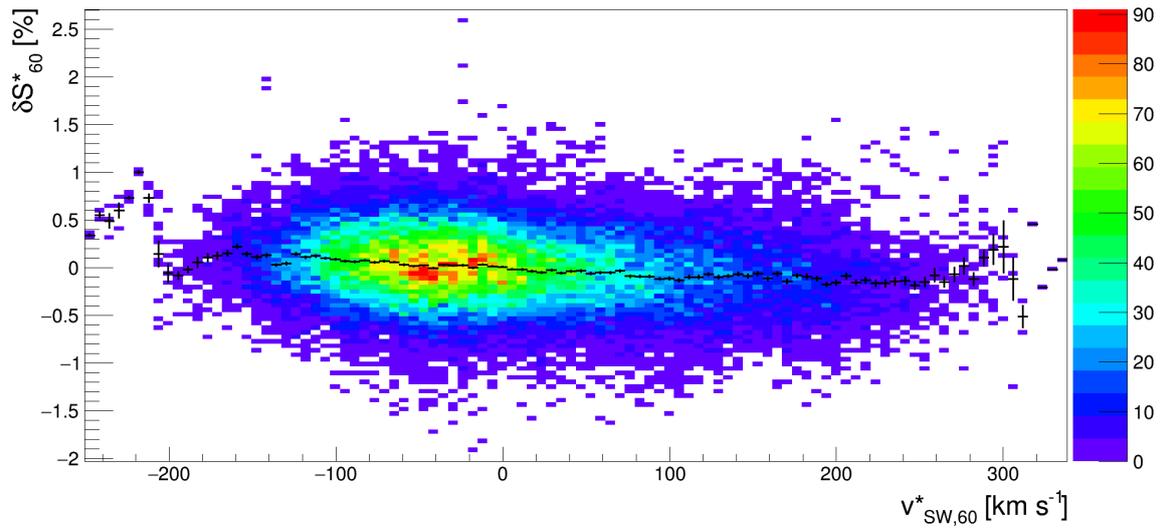
Figure A.16: Variation in Scaler rate (red points in upper graph) and Solar Wind velocity (blue points in lower graph) after subtracting the respective centered moving average data for a δt of (a) 75 min and (b) 90 min.

Correlation of Auger Scaler rate with Solar wind velocity



(a) $\delta t = 45$ min

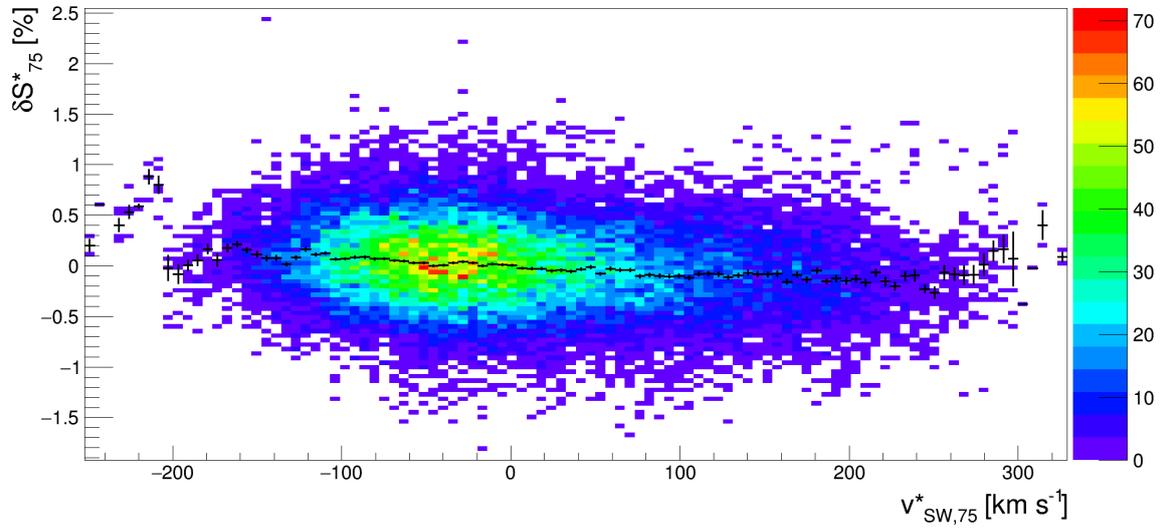
Correlation of Auger Scaler rate with Solar wind velocity



(b) $\delta t = 60$ min

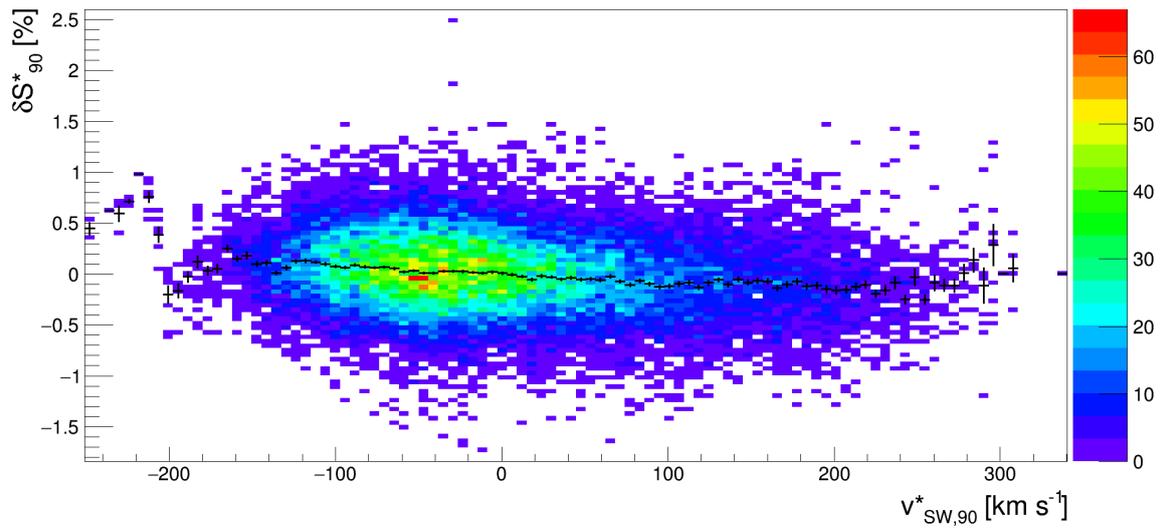
Figure A.17: 2D-histogram of $\delta S^*_{\delta t}$ and $v^*_{SW,\delta t}$ with profile line (black) for each bin in $v^*_{SW,\delta t}$ for a δt of (a) 45 min and (b) 60 min.

Correlation of Auger Scaler rate with Solar wind velocity



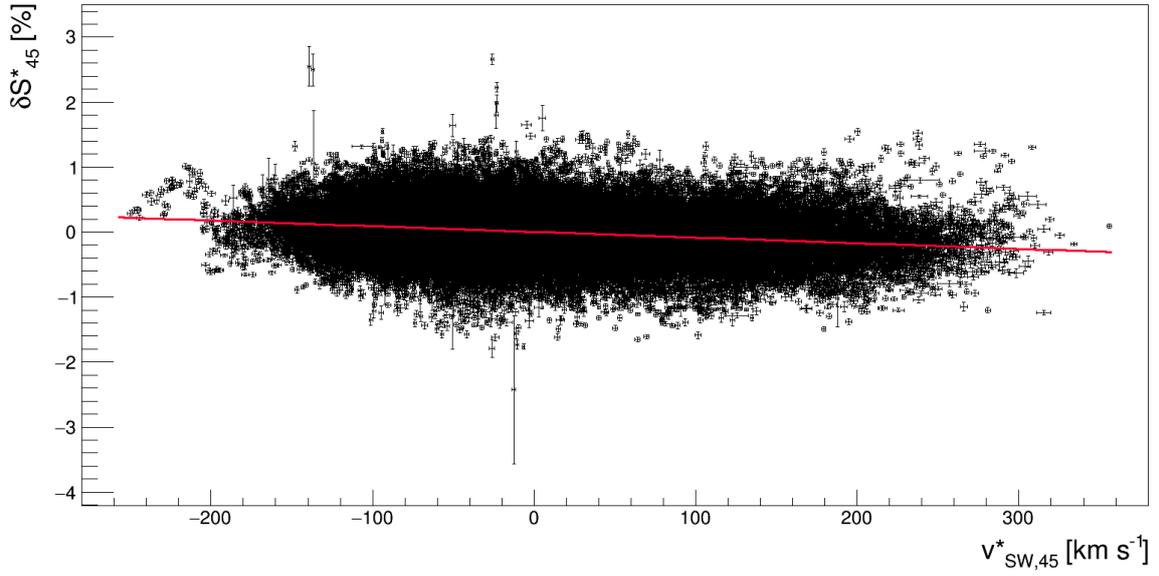
(a) $\delta t = 75$ min

Correlation of Auger Scaler rate with Solar wind velocity

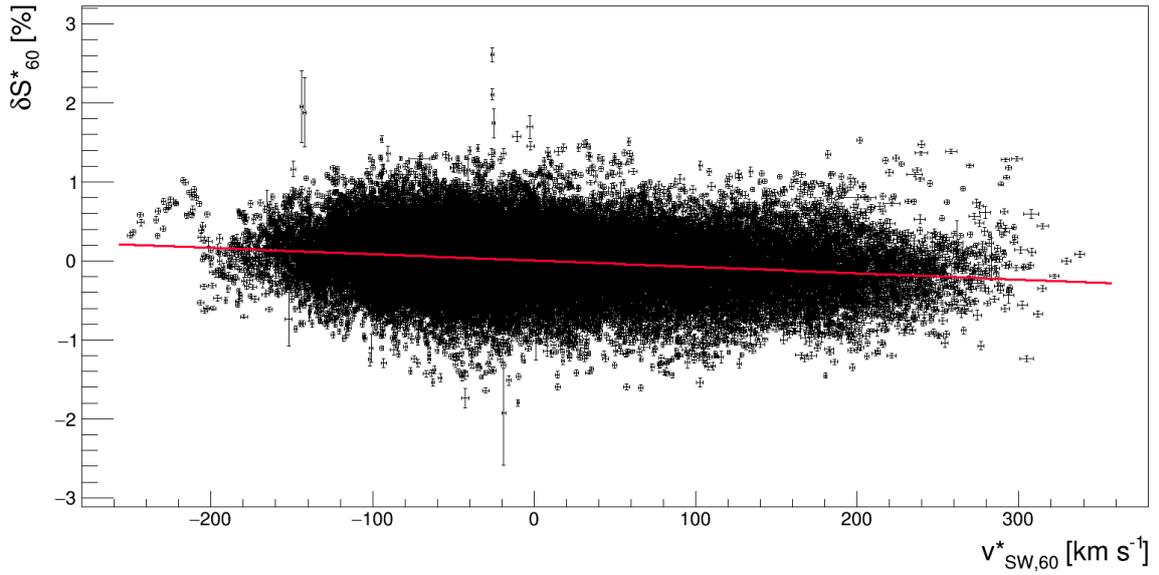


(b) $\delta t = 90$ min

Figure A.18: 2D-histogram of $\delta S^*_{\delta t}$ and $v^*_{SW,\delta t}$ with profile line (black) for each bin in $v^*_{SW,\delta t}$ for a δt of (a) 75 min and (b) 90 min.

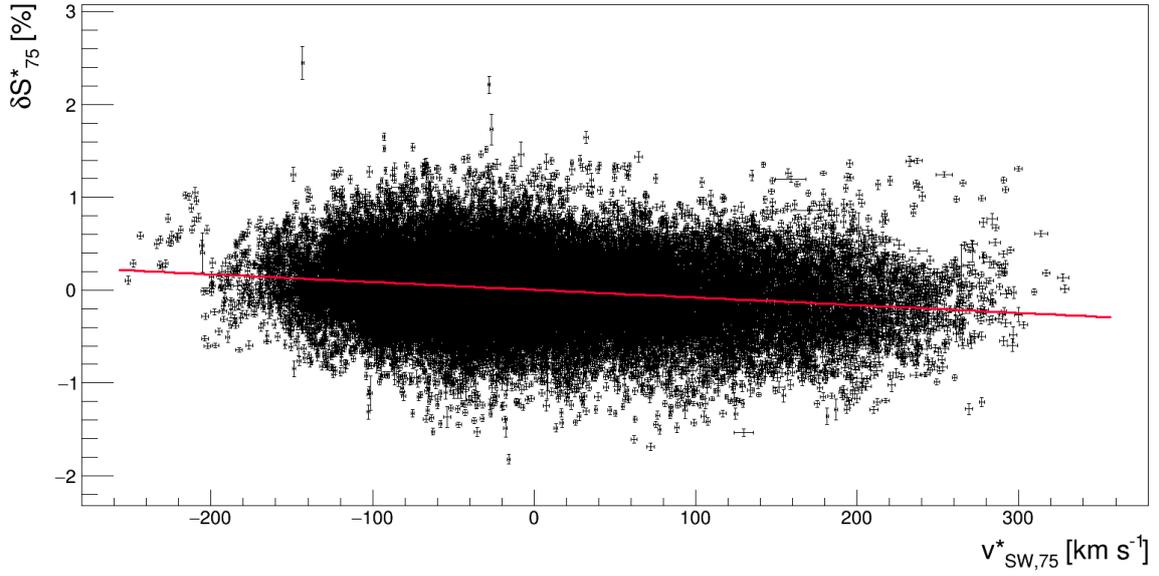


(a) $\delta t = 45$ min

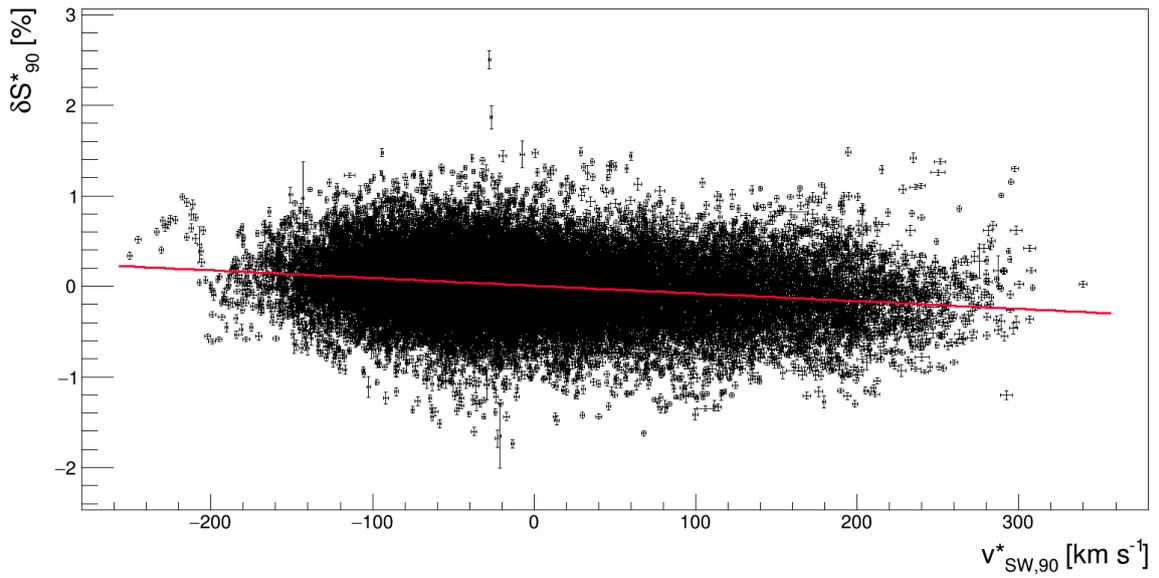


(b) $\delta t = 60$ min

Figure A.19: Correlation of $\delta S_{\delta t}^*$ and $v_{SW,\delta t}^*$ with linear fit (red line) for a δt of (a) 45 min and (b) 60 min.

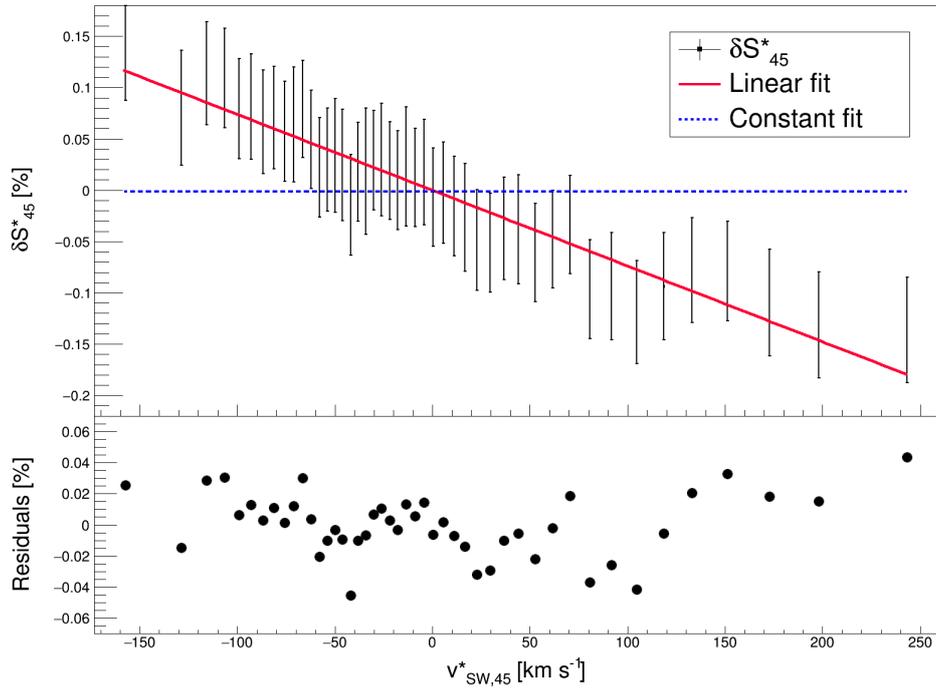


(a) $\delta t = 75$ min

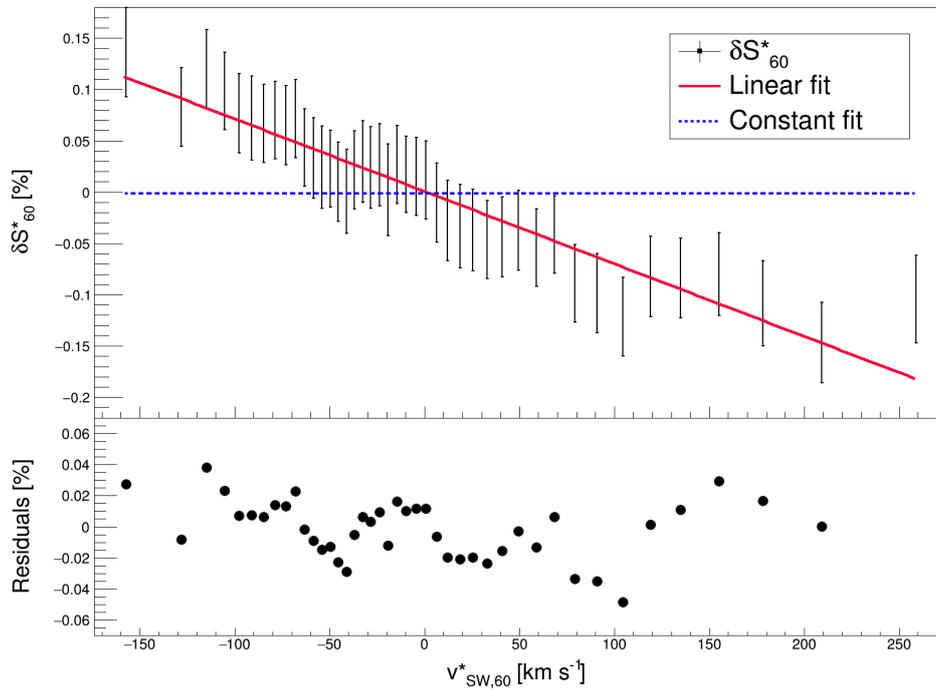


(b) $\delta t = 90$ min

Figure A.20: Correlation of $\delta S_{\delta t}^*$ and $v_{SW,\delta t}^*$ with linear fit (red line) for a δt of (a) 75 min and (b) 90 min.

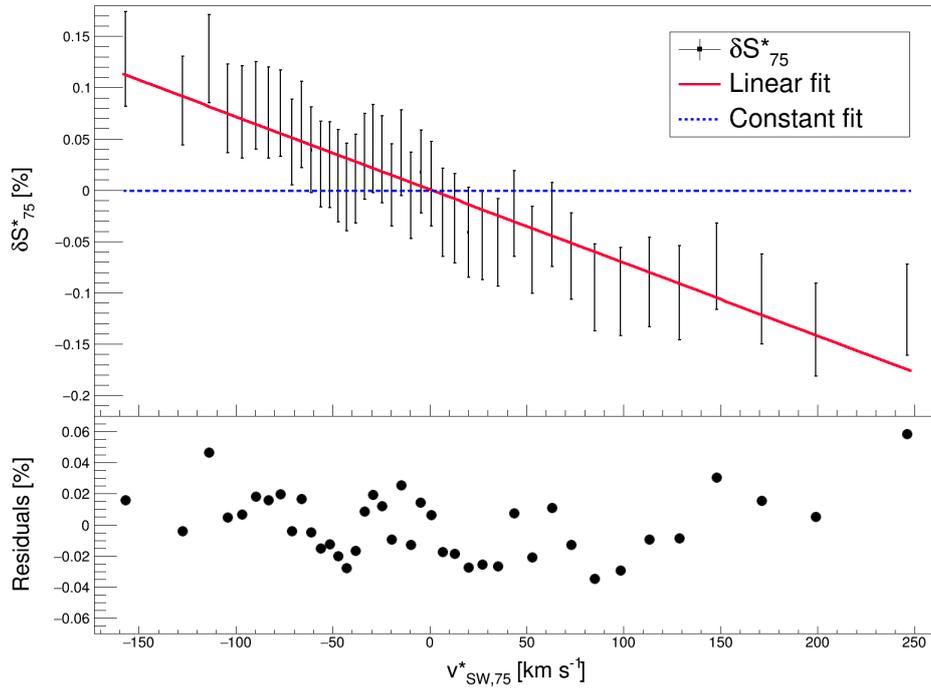


(a) $\delta t = 45$ min

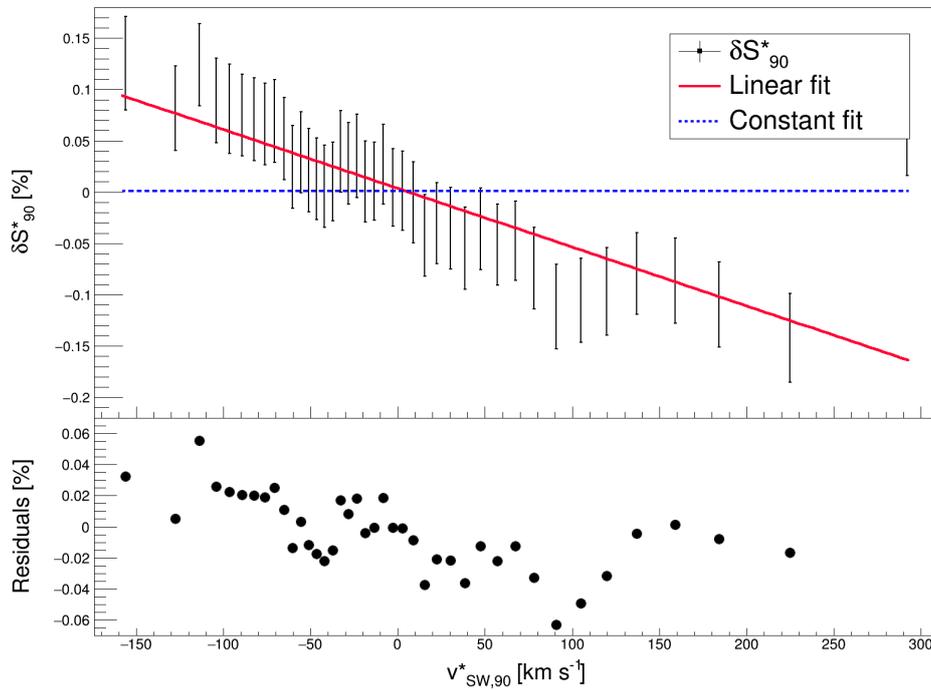


(b) $\delta t = 60$ min

Figure A.21: Correlation of $\delta S_{\delta t}^*$ and $v_{SW,\delta t}^*$ with linear fit (red line) and constant fit (blue dotted line) for a δt of (a) 45 min and (b) 60 min. Residuals (black dots) are shown for the linear fit.



(a) $\delta t = 75$ min



(b) $\delta t = 90$ min

Figure A.22: Correlation of $\delta S^*_{\delta t}$ and $v^*_{SW,\delta t}$ with linear fit (red line) and constant fit (blue dotted line) for a δt of (a) 75 min and (b) 90 min. Residuals (black dots) are shown for the linear fit.

References

- [1] H. Kojima et al., *Dependence of cosmic ray intensity on variation of solar wind velocity measured by the GRAPES-3 experiment for space weather studies*, *Phys. Rev. D.* **91**, 121303, (2015).
- [2] W. Kolhörster, I. Matthes and E. Weber, *Gekoppelte Höhenstrahlen*, *Naturwissenschaften* 26: 576, 1938.
- [3] Pierre Auger et al., *Extensive Cosmic-Ray Showers*, *Rev. Mod. Phys.* **11**, 288 (1939).
- [4] CERN – Restarting the LHC at 13 TeV – <https://home.cern/about/engineering/restarting-lhc-why-13-tev> , last visited 26.09.2017.
- [5] L. Dorman, *Cosmic ray interactions, propagation, and acceleration in space plasmas*. Springer, 2006.
- [6] W. Bietenholz, *The most powerful particles in the Universe: a cosmic smash* (2013), [arXiv:physics.hist-ph/1305.1346].
- [7] A. Haungs et al., *KCDC - The KASCADE Cosmic-ray Data Centre*, *J. Phys. Conf. Ser.* **632**, No. 1, 012011 (2015), [arXiv:astro-ph.IM/1504.06696].
- [8] Wikimedia Commons – Pierre Auger Observatory – https://commons.wikimedia.org/wiki/File:Layout_of_Pierre_Auger_Observatory.svg , last visited 26.09.2017.
- [9] Pierre Auger Observatory – Surface Detector station – <https://www.auger.org/index.php/observatory/auger-hybrid-detector> , last visited 26.09.2017.
- [10] B. Carroll, D. Ostlie, *An Introduction to Modern Astrophysics*, Benjamin Cummings, ISBN 0-201-54730-9, (1995).
- [11] Wikimedia Commons – Structure of the magnetosphere – https://commons.wikimedia.org/wiki/File:Structure_of_the_magnetosphere-en.svg , last visited 26.09.2017.
- [12] J. J. Masias-Meza, *Solar Cycle Modulation of Cosmic Rays Observed with the Low Energy Modes of the Pierre Auger Observatory*, *Proc. ICRC* (2015), **142**.

References

- [13] The Pierre Auger Collaboration, *The Pierre Auger Observatory scaler mode for the study of solar activity modulation of galactic cosmic rays*, *JINST*, **6**, P01003, 2011.
- [14] R. A. Matzner, *Dictionary of Geophysics, Astrophysics and Astronomy*. CRC Press, 2001.
- [15] E.L. Cussler, *Diffusion. Mass Transfer in Fluid Systems*. Cambridge University Press, 1997.
- [16] X. Bertou, D. Allard, *Current status of the scalers: Towards the detection of GRBs with Auger?*, GAP2005_053, Internal reports for the Pierre Auger Collaboration, 2005.
- [17] X. Bertou, *Analysis of the Auger scaler data in search for GRBs*, GAP2006_100, Internal reports for the Pierre Auger Collaboration, 2006.
- [18] Space Weather page – Scaler data – <http://auger.colostate.edu/ED/scaler.php>, last visited 26.09.2017.
- [19] H. Asorey, E. Roulet, *The new weather data sets for the Auger Observatory site*, GAP2016_049, Internal reports for the Pierre Auger Collaboration, 2016.
- [20] NASA OMNIWeb – High resolution solar wind data – https://omniweb.gsfc.nasa.gov/ow_min.html , last visited 26.09.2017.
- [21] Homepage of ACE – <http://www.srl.caltech.edu/ACE/> , last visited 26.09.2017.
- [22] Homepage of GEOTAIL – <http://www.stp.isas.jaxa.jp/geotail/> , last visited 26.09.2017.
- [23] Homepage of WIND – <https://wind.nasa.gov/> , last visited 26.09.2017.
- [24] NASA OMNIWeb – Data description – https://omniweb.gsfc.nasa.gov/html/omni_min_data.html , last visited 26.09.2017.
- [25] J. C. Kasper et al., *Physics-based tests to identify the accuracy of solar wind ion measurements: A case study with the Wind Faraday Cups*, *J. Geophys. Res.* **111** (2006), no. A3 2156-2202.
- [26] OMNIWeb Data Documentation – Data cleaning – https://omniweb.gsfc.nasa.gov/html/ow_data.html#data_clean, last visited 26.09.2017.
- [27] Sonnerup, B.U.O and L.J. Cahill, *Explorer 12 observations of the magnetopause current layer*, *J. Geophys. Res.* **73**, 1757, 1968.
- [28] Kiel neutron monitor data – <http://134.245.132.179/kiel/main.htm> , last visited 26.09.2017.
- [29] McMurdo neutron monitor data – http://neutronm.bartol.udel.edu/~pyle/bri_table.html , last visited 26.09.2017.

References

- [30] A. Khalil, *Studies of astrophysical very-high energy gamma-ray emission with the Pierre Auger Observatory*, GAP2017_013, Internal reports for the Pierre Auger Collaboration, 2017.
- [31] T3 Bad Periods – <http://ipnwww.in2p3.fr/~augers/AugerProtected/ICRC.html> , last visited 26.09.2017.
- [32] H. Asorey, X. Bertou, *First large timescale analysis of Auger SD scaler data: Towards cosmic ray Solar modulation studies*, GAP2008_072, Internal reports for the Pierre Auger Collaboration, 2008.

Erklärung

Hiermit erkläre ich, Daniel Hans-Gerd Axel Rosenbaum, dass ich diese Arbeit zum Thema

Correlation of Solar Wind velocity and Scaler data of the Pierre Auger Observatory

selbstständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate kenntlich gemacht habe.

Wuppertal, den 26. September 2017

(Daniel Hans-Gerd Axel Rosenbaum)
