Optimal signal time for the directional reconstruction of Cosmic Rays using radio detection at the Pierre Auger Observatory

Alexander van Ratingen

July 12, 2011 - August 24, 2012

Abstract

The Pierre Auger Observatory's goal is to determine the nature and origin of ultra high energy cosmic rays. To reconstruct the direction of the cosmic ray, one of the most important parameters is the arrival time of the shower at the detector. For the new radio detection setups, the shower arrival time is measured as a pulse in the electric field. While the arrival time is currently defined as the time when the maximum of that pulse is recorded, this thesis investigates the use of the pulse start for the directional reconstruction.

Despite theoretical advantages and thus an expected improvement, this thesis shows that the current arrival time reconstruction method cannot be improved by basing it on the pulse start time instead of the pulse maximum using the presented algorithm.

Contents

1 Introduction

This thesis is the result of research for a bachelor internship at the experimental high energy physics department at the Radboud University Nijmegen. In my time there, I worked on improving upon part of the detection of ultra high energy cosmic rays (UHECRs).

I was supervised by Charles Timmermans, and received a lot of help from Harm Schoorlemmer and Stefan Grebe. Gratefully, I'd like to thank them for their patience and assistance.

1.1 Cosmic Rays

Cosmic rays were first observed in the beginning of the twentieth century as ionizing radiation in the atmosphere. In 1912 by Victor Hess made a balloon flight with detectors and showed that the amount of radiation increased with altitude. From this, he concluded that cosmic rays originate in space. This discovery won him the Nobel prize. In the following years, research was devoted to the nature of cosmic rays. The experiments conducted showed the radiation can be highly energetic and consists mostly of protons.

When a cosmic ray enters the Earth's atmosphere, the resulting collisions create a shower of secondary particles. This was proven by Pierre Auger by placing two particle detectors separated by some distance, and showing that they measure events simultaneously. This shower of particles consists of billions of particles and spreads several kilometers.

The research of UHECRs was started by John Linsley who was the first to discover a particle with a macroscopic energy of 10^{20} eV ($\sim$ 16 J) in 1962.[6] The highest energy cosmic rays are very rare, from one event square kilometer per year to one per square kilometer per century. Image 1 shows the flux of cosmic rays for the complete energy spectrum. This end of the spectrum is interesting, because it is currently unknown what processes can lead to such high energies.

Like all charged particles, these UHECRs are deflected by the magnetic fields that exist in space. However, because of their ultra high energy, that deflection is only small. Thus, the direction from which they arrive points back to their origin.

Another interaction of these particles is with the cosmic microwave background (CMB). The most important of these is the Greisen-Zatsepin-Kuzmin (GZK) effect; a cosmic ray interacts with a CMB photon and creates a pion. This process leads to a significant loss of energy; it puts a limit on the distance an UHECR can propagate before having lost most of its energy. There is only a 0.1 % chance for a $3 \cdot 10^{20}$ eV cosmic ray to have travelled more than 50 Mpc.[2, 1]

Thus, UHECRs are interesting because their arrival direction point back to their origin which must lie somewhere in our cosmic neighbourhood ($\leq$ 200 Mpc), and that place of origin is of particular interest because it may lead to a new understanding of physics.

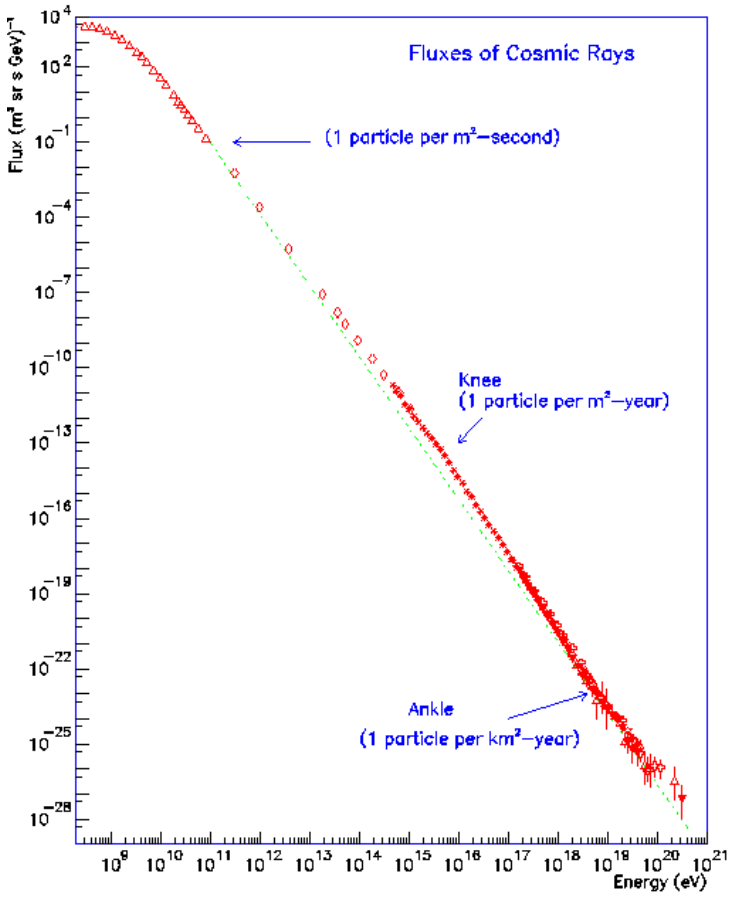

Figure 1: Flux of cosmic rays.

2 Detection methods

Cosmic ray detection took off in the 1960s, though progress was very slow. Many techniques to detect shower properties were developed over time. Among the most successful were particle and fluorescence detection. But also radio detection was developed. [4]

2.1 Particle detection

Particle detectors measure the part of the shower that reaches the earth's surface. These detectors simply measure the amount of radiation left to deduce properties like the energy of the primary particle. However, at sea level, the shower has passed it maximum size and is already decreasing. Thus, this method of detection provides only a limited view into the shower.

2.2 Radio detection

Radio detection is becoming a more popular method of investigating a cosmic ray shower. Several different processes in the shower emit electromagnetic radiation that can be detected. The Cherenkov radiation that is produced in the atmosphere was one of the first detected effects. Despite being quite a technical challenge, this method was successfully used in the 1960s.

Processes that produce detectable radiation originate from the negative charge excess of shower, from a geomagnetically induced electric dipole arising from Lorenz separation of positive and negative charge and from the geomagnetically induced transverse current within the showerfront. The latter is the most dominant process for almost all distances and frequencies. Charge excess is about an order of magnitude less efficient, whereas the dipole induced by the charge separation is only important at high frequencies close to the shower core.

Initially, there were problems with triggering the radio detection readout when a cosmic ray event occurs. For this reason, radio detection was always triggered by particle detection. The development of truly stand-alone radio detection wasn't resumed until the turn of the century.

As technical advances were made in electronic hardware, the world became increasingly less radio-quiet. Cherenkov detection went into a slow decline through the 1970s, and there was little progress on other methods of cosmic ray detection. A revived interest in cosmic ray detection and an increase in research led to further development of the field around the year 2000. [4]

3 The Pierre Auger Observatory

The Pierre Auger Observatory researches UHECRs with the ultimate goal to determine their nature and origin, learn what processes could accelerate particles to such high energies and answer many other open questions. It's the largest cosmic ray array ever built, as to maximize the exposure to the very rare highest energy cosmic rays. First proposed in 1992, it's been collecting data since the year 2004.

It currently offers three detection methods to observe the showers created by cosmic ray collisions with the earth's atmosphere. Besides the established particle and fluorescence detectors, efforts are made to include radio antennas as a third independent detection method to complement the existing techniques. The main advantage is its high duty cycle; nearly 100%, compared to about 10% for the existing hybrid method due to the fact that fluorescence detection is only available during nighttime in the right weather conditions.

3.1 Surface detection array

The Auger surface detector array consists of 1660 stations over a 3000 km^2 triangular grid. Each station has a tank that contains 12000 l of pure water, read out by three photomultiplier tubes (PMTs). The secondary shower particles passing through the water tank emit Cherenkov radiation, which is detected by the PMTs. The resulting data is used to reconstruct the trigger time and signal strength, and determine the shower parameters.

3.2 The Auger Engineering Radio Array

For radio detection, the new Auger Engineering Radio Array (AERA) setup provides significant improvements over its predecessors. It became operational about halfway during this internship in April 2011. Using 21 antennas[1] to detect the electromagnetic field at ground level, traces of radio emission from the airshowers are recorded.

These stations are laid out in a triangular grid, with 150 m spacing between them, making the array sensitive to showers with energies of 10^{17} eV and up. Their antennas are sensitive to signals in the band between 27 MHz and 84 MHz, chosen for being relatively radio-quiet and well within the band of coherent shower frequencies that ranges from 10 to 100 MHz. Both the north-south and east-west polarization components of the radio signal are measured, amplified, filtered and then fed to a 12-bit 200 MHz digitizer. When triggered, a trace consisting of 2048 voltage samples is stored for analysis.[5]

From this data, many of the showers parameters can be reconstructed. The most important parameter for the directional reconstruction is the arrival time of the pulse. These arrival times of the pulse at different stations can be used in a plane fit — or a more elaborated method that takes into account the showerfront curvature — which directly determines the shower direction and thus the direction of the primary cosmic ray.

To reconstruct the shower parameters from detector data, several software packages exists. The one most widely used in the Auger collaboration is the Offline framework, which also has support for radio detection. The current directional reconstruction for radio data is based on a plane fit and agrees up to $\mathcal{O}(10^\circ)$ with the established detection methods, as can be concluded from figure 2.

4 Analysis

During my internship, I worked with the Offline framework for reconstructing shower parameters from data measured by the AERA setup at the Pierre Auger Observatory.

4.1 Goal

In the current implementation of the Offline framework, the time at which a radio event is defined to occur, the *event station arrival time*, is by the instance of time at which the pulse in the measured data attains its maximum[2]. These timings for all stations that measured the event are then used in a fit to the showerfront shape to obtain parameters like the shower's main axis, or its radius of curvature. Since the shower's axis directly translates to the primary particles trajectory and thus to its origin, it is one of the most important parameters. The station arrival time's importance to this reconstruction makes it a good target for possible improvements. The goal of this analysis is exactly this; an attempt at an improvement of the station arrival time determination method used by Offline.

[1] The intention is to increase the coverage to 20 km^2 with a total of 162 antennas.

[2] After suitable preparation of the data, and using some function of the three-dimensional electromagnetic field trace. This is elaborated further in 4.2.1.

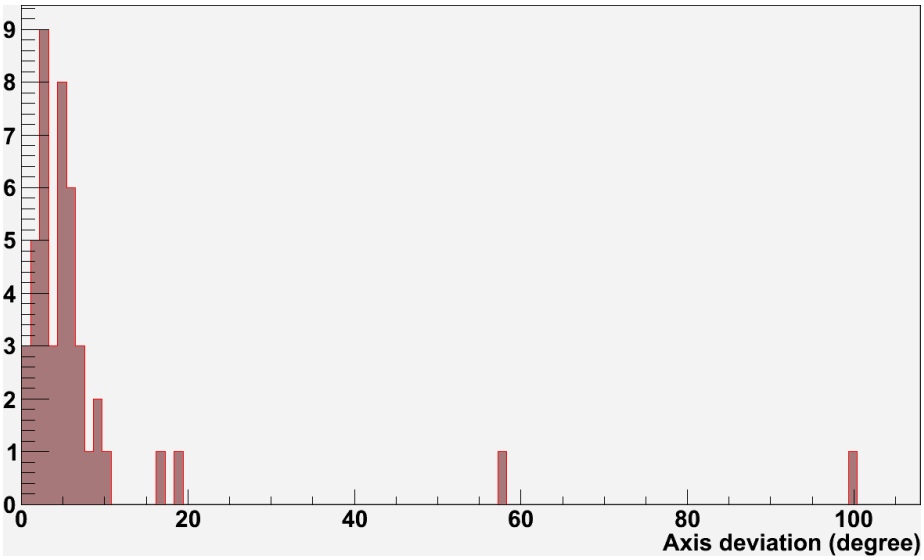

Figure 2: Distribution of angles between reconstructions of the shower axis using the hybrid setup and the AERA array. Results using the default Offline reconstruction as described in 4.2.1. Events with very high deviations ($> 45°$) have low SNR values.

The current definition of the arrival time is clearly defined for trivial pulses and still agreeable for realistic ones, but one may wonder whether it's the most optimal. Alternatively, the pulse start seems to be a more natural definition of the arrival time. Because the pulse width is not constant within events, the two lead to different results. The pulse start suffers from the disadvantage over the maximum that it is much less distinguishable in a noisy signal. It is therefore not a priori clear that a pulse-start based method will improve results.

Our attempt at improving upon part of the reconstruction will focus on finding a suitable definition of the pulse start, implementing it, and verifying its results.

Since the final shower axis value is very sensitive to the arrival time reconstruction, it is a good indicator of the performance of that reconstruction. Thus, in the following we will use deviations of the shower axis as proportional to the error in the arrival time method.

We'll substitute the shower axis as reconstructed using hybrid data for the actual axis deviation in this comparison. Because of this, uncertainties of the results will be at least as great as those of the hybrid reconstruction. This lower limit is $1.2°$ for our dataset.[3]

4.2 Method

To attempt to optimize the shower reconstruction, a new algorithm is created that uses the pulse start instead of its maximum for the arrival time reconstruction. First we describe the current process, since it's the base for all changes and the standard we're comparing results with.

4.2.1 Current reconstruction

The electric field After the raw measured data has been fed into the framework's processing loop, the first order of business is to remove any detector influence. The raw data holds time series of digital counts for all stations included in the event. For each station two horizontal components of the electric field are recorded. These are first converted to voltages using implementation characteristics, such as minimum and maximum voltage that can be measured and bit depth of the analogue to digital converter.

Any DC-offset is removed to obtain the measured voltage in the antenna from radiation. At this point, the frequency dependent gain of the antenna, and any influence of the electronics on the phase of the signal is removed to obtain a result that resembles the electric field as it existed during the measurement by the antenna.[3]

Next up is removing any noise from the signal caused by anything other than the airshower. A bandpass filter removes any signal outside of the frequency band our antenna is most sensitive to or is known to be uncorrelated with the airshower signal. To minimize the negative effects of this filter at the borders of the frequency band we apply a windowing function that approaches zero at the borders. To minimize narrow band noise, a median filter is applied over a sliding window of the time series signal. Finally, the direction dependent sensitivity is removed. Knowing the antenna response pattern, the electric field can be obtained from the measured field.

Since the antenna only measures the field in two orthogonal directions, determining the full vector field requires one more piece of information. For this, the incident direction of the radiation is used. Knowing that the radiation is a transverse wave, the electric field vector must be orthogonal to the direction of propagation. Since the direction is unknown because it depends on on the data we're still reconstructing, we start with an approximation of the direction and iteratively reconstruct it with increasingly better directional knowledge.

Finally, knowing the antenna response pattern, the electric field is obtained from the measured field by removing the direction dependent sensitivity. Note that these last steps make the assumption that there is only radiation from one direction, anything emitted from other sources is incorrectly adjusted for. This generally does not matter since the signal of the shower event is dominant.

Improving the signal Instead of working with the electric field itself, it can be more useful to deal with a transformed signal. The Offline framework offers several built in transformations that can be applied to the signal. Naturally, the Fourier transform is one of them; it's used in various places throughout the standard reconstruction pipeline. However, from the frequency domain data that the Fourier transform gives us, we always return to the time domain. Obviously, we need to when we try to determine the arrival time.

For our purposes, we'll briefly discuss another useful signal transformation that is applied to obtain a new but similar signal that is still in the time domain; the Hilbert envelope. Informally, this transformation removes any phase

[3] The only structural error in obtaining the real electric field we haven't corrected for by now is the directional dependence. We do this later on because we need need to reconstruct the direction first.

information leaving only the amplitude, by averaging out the signal. The result is a positive signal.

To compute it, we first take the Hilbert transform of the signal, loosely[4] defined as:

$$f(t) \quad \xrightarrow{\mathcal{H}} \quad \mathcal{H}\{f(t)\} \;=\; \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} d\tau \tag{1}$$

This illustrates the averaging effect, the new value at t is determined by the value of the original value an all points in time τ weighted by the inverse of the distance $(t-\tau)$ to that point. Because of the asymmetry in the weighting factor, it changes the phase of the components depending on the sign of their frequency, and leaves the amplitudes in place. This effect is evident from the effect of the Fourier transform on a Hilbert transformed function.

$$f(t) \quad \xrightarrow{\mathcal{H}} \quad \mathcal{H}\{f(t)\} \quad \xrightarrow{\mathcal{F}} \quad [\mathcal{F}\cdot\mathcal{H}]\;\{f(t)\} = -i\,\mathrm{sign}(\omega)\,\mathcal{F}\{f(t)\} \tag{2}$$

In words, the Fourier transform of a Hilbert transformed function is equal to just the Fourier transform multiplied with $\pm i$ depending on the sign of the frequency, so rotated by $\pm\frac{\pi}{2}$ in the complex plane (phase space).

The Hilbert envelope is then simply given by

$$\sqrt{f(t)^2 + \mathcal{H}\{f(t)\}^2} \tag{3}$$

The resulting function has some nice features such as being positive and greater-than-or-equal-to the original function. Also, the envelope of a sine function is unity.[5]

Applying this transformation is optional in Offline, but it generally improves the reconstruction. Therefore we'll always include it in the default reconstruction we're comparing improvements with.

The signal arrival time Now that we've approximated the electric field in the stations that measured the shower, we can determine the signal arrival time. For this, the vector field trace is converted to a scalar trace by taking either the absolute value or selecting a component. Then, the signal arrival time is taken to be the moment in the trace where the maximum value is first obtained.

The shower parameters With the data we have now, all shower parameters can be deduced. We'll restrict ourselves to the shower axis since that's the one we're interested in.

From the arrival times of the signal at different stations, we can fit the parameters that describe the showerfront. For each one of these parameters, you need one single piece of information to be able to reconstruct. That information comes in the form of station arrival times. A simple flat showerfront has three parameters: two for the orientation, and one for the timing. Thus, reconstructing a plane fit requires arrival times available for at least three stations. Currently, Offline only supports a plane fit for determining the shower axis. However, efforts are put into including a more sophisticated showerfront fit with additional parameters, showerfront curvature for one.

[4] Because of the singularity of the integrand at $t=\tau$, the Cauchy principal value must be taken: split the integral in two parts that approach the singularity.

[5] The Hilbert transform of a sine is minus a cosine, and summing them squared is unity. Thus, the Hilbert envelope of a sine is unity.

4.2.2 Tested improvement

The main idea for an optimization of the above outlined reconstruction comes from the fact that the signal at the pulse start is a much more natural definition of the arrival time. As such, it does not include an error from the variation in pulse shapes for different stations like the pulse maximum does. The main error that is introduced in the original implementation is due to a variation in pulse duration at different distances from the shower core.

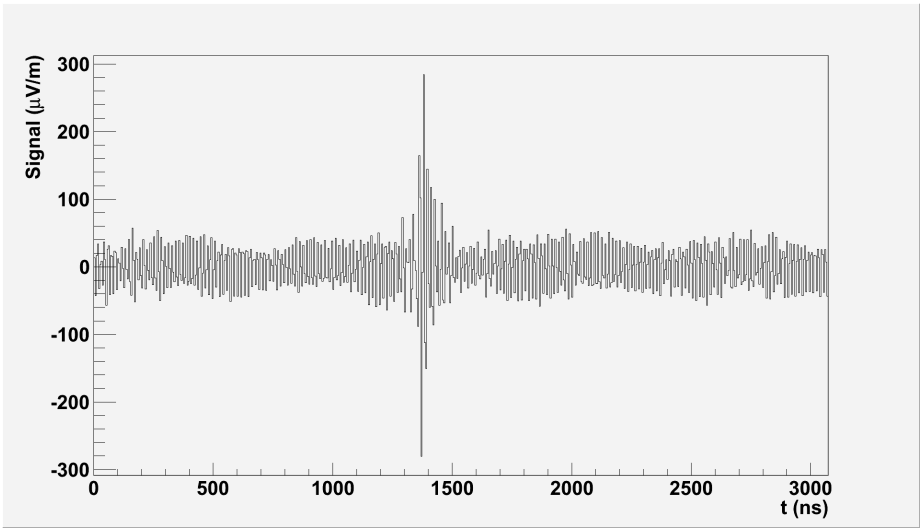

Figure 3: A signal trace we'll follow as it is processed.

The optimized reconstruction replaces the current implementation of the method by which the station data is combined into event data. At this point, the signal used is the preprocessed signal trace which has been passed through a Hilbert envelope. This is still the same as it would be in the original reconstruction. Figure 3 shows an actual trace before the Hilbert envelope is applied, figure 4 shows the resulting evelope. The next step is where we deviate.

First, the signal in the time domain is integrated, resulting in a continuously increasing trace.

$$I(t) = \int_{t_0}^{t} S(\tau)\, d\tau$$

In the absence of a pulse, the signal is only subject to background noise. The integrated signal, then, is still an increasing function where the overall slope is an indicator of the noise level. This can clearly be seen in figure 5. If we fit a linear function to this data, we can subtract the constant noise n.

$$M(t) = I(t) - nt$$

We arrive at a trace that is no longer increasing at all times, but roughly constant in the absence of any pulses. Image 6 shows the result of the process. The signal arrival time t_a is defined as the moment in time where the integrated

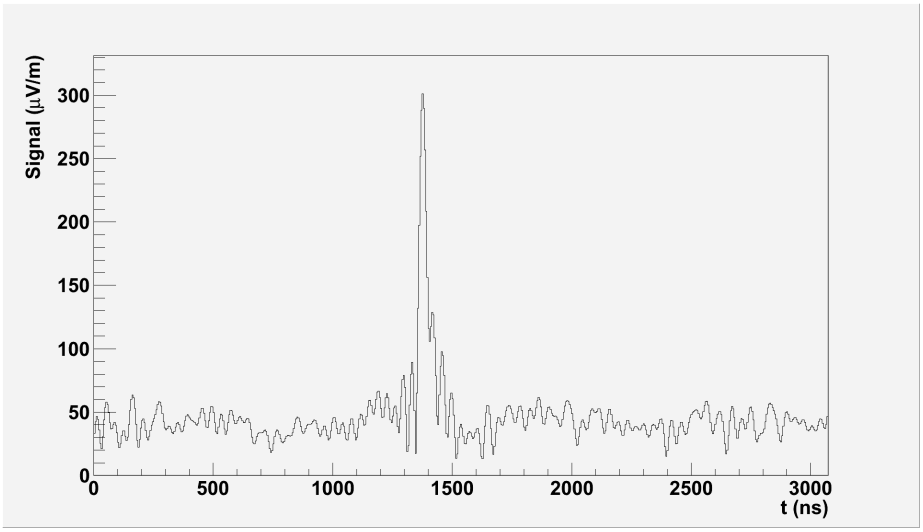

Figure 4: The signal trace with the Hilbert envelope applied.

trace first reaches a certain fraction T of the average final trace value m.

$$\frac{M(t_a)}{m} = T$$

We'll refer to this fraction as the *threshold* defining the pulse arrival time at the station.

Since the overall trend in the trace during the pulse is increasing, a low threshold value results in an arrival time at the start of the pulse. For symmetric pulses[6], a threshold value of $\frac{1}{2}$ would result in the same arrival time as a method that uses the pulse maximum.

4.3 Results

The radio event data used was taken in the period between April 2011 and January 2012, since it contains coincident events of both the hybrid and the latest AERA radio setup. This set consists of 45 events. Because of this small dataset, no data quality cuts were made.

Applying the reconstruction to this complete set of events for the original arrival time method and the modified version using various values of the threshold parameter, we obtain the shower axes for the events. These are compared with the axes as reconstructed using hybrid data in figure 7.

Comparing the two methods, we see that they yield similar results, indicating that they are both at the very least reasonable approaches to determining the signal arrival time. However, the angles from the hybrid shower axis are clearly smaller for the original algorithm. The average angle as well as the deviation have both increased in using the modified method for all threshold values.

[6]Radio events are asymmetric in this respect, they have longer tails.

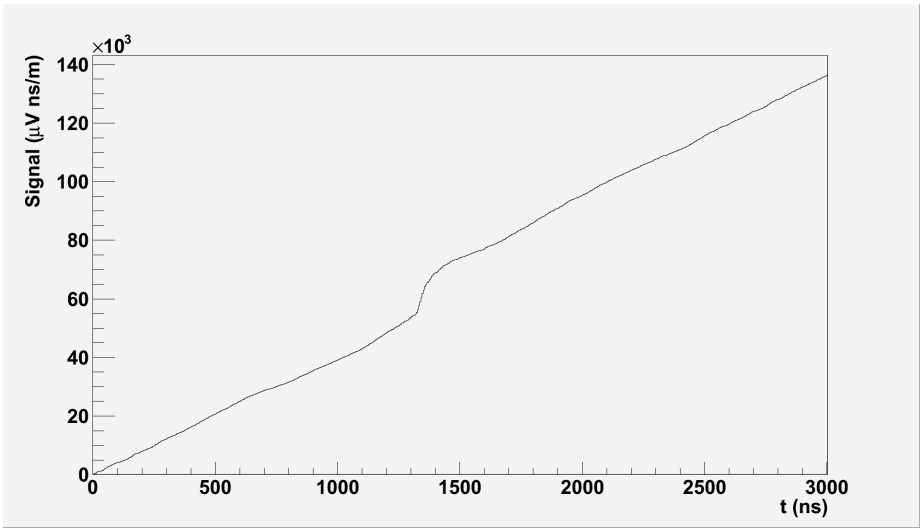

Figure 5: The signal trace with the Hilbert envelope applied, and after it's been integrated. One can clearly distinguish the pulse from the background by the significant change in slope.

Threshold	Mean (°)	Std. deviation (°)
original	8.23	6.9 ± 1.3
0.05	11.51	14.3 ± 3.4
0.10	11.51	14.3 ± 3.4
0.15	11.49	13.4 ± 3.0
0.20	11.51	12.4 ± 3.3
0.25	11.51	12.4 ± 3.3
0.30	11.52	12.4 ± 3.3
0.35	11.63	14.3 ± 3.3
0.40	11.61	15.4 ± 3.7
0.45	11.44	14.8 ± 3.8
0.50	11.44	14.8 ± 3.8
0.55	11.44	14.8 ± 3.8
0.60	11.42	14.8 ± 3.8
0.65	11.44	14.8 ± 3.8
0.70	11.44	17.5 ± 5.0
0.75	11.38	13.5 ± 3.5
0.80	11.38	13.5 ± 3.5
0.85	11.31	13.2 ± 3.5
0.90	11.39	13.3 ± 3.4
0.95	11.33	11.5 ± 2.6

Table 1: Results for varying threshold values. Standard deviations are obtained by fitting the histogram of angles α between the hybrid and the radio reconstruction with a gaussian-like function: α gaus(α).

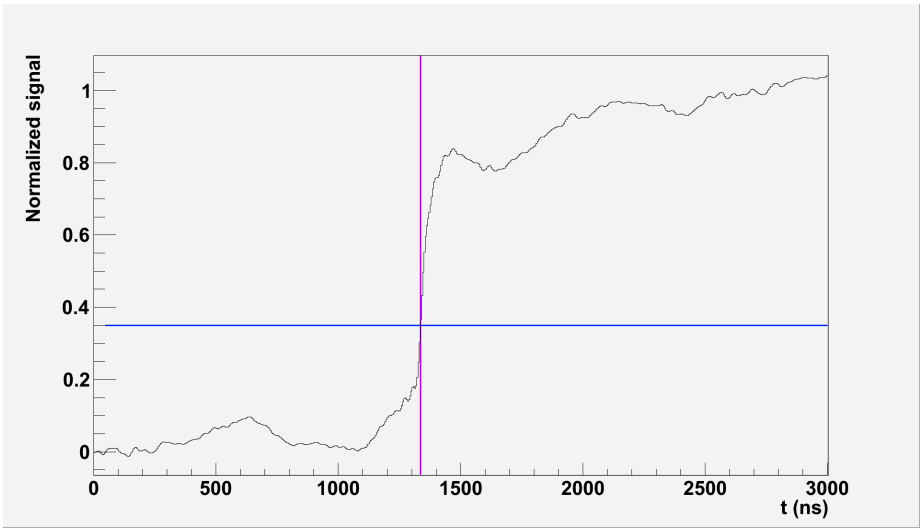

Figure 6: The final normalized trace after processing. This is where we intersect the trace with a threshold value (horizontal) to find the arrival time (vertical).

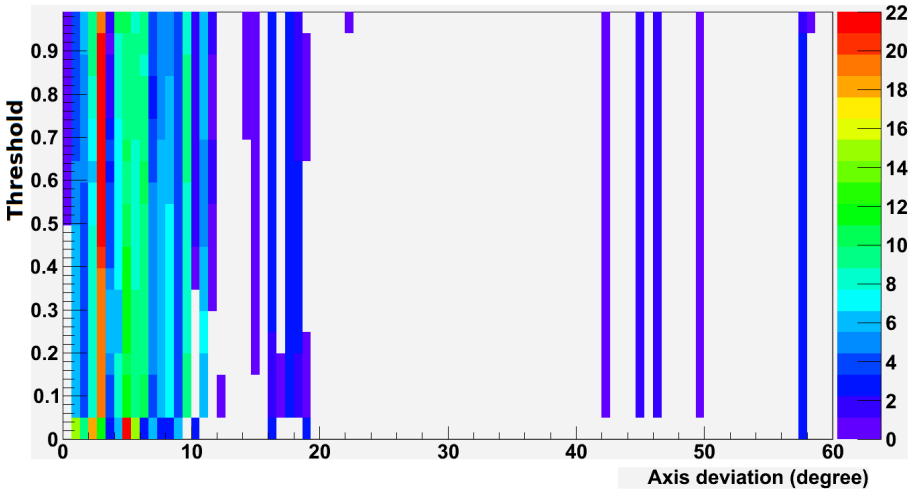

Figure 7: Distribution of angles between shower axes as reconstructed using radio data and hybrid data as a reference, for various values of the threshold parameter of the arrival time algorithm. A threshold value of zero indicates the original reconstruction.

These results also reveal that the actual value of the threshold parameter does not have much effect on the reconstruction outcome for the shower axis. This means that, for all stations within an event, the normalized traces that are intersected as the last step of the reconstruction process, are actually quite similar in shape. Thus, a different threshold value will result in a station wide

constant offset, resulting in nearly the same shower axis. Figure 8 demonstrates this, by plotting these normalized traces for all stations in all events.

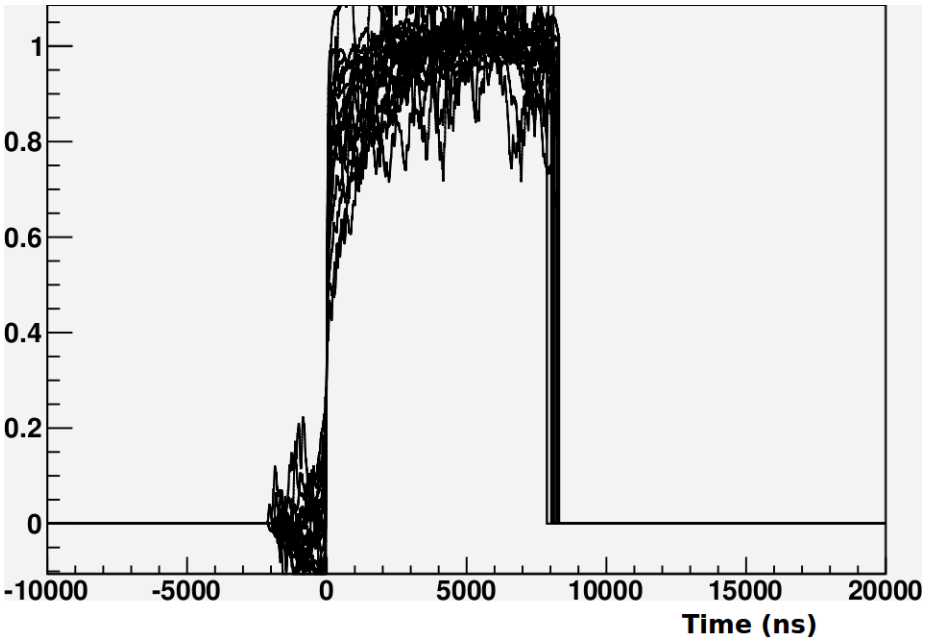

Figure 8: Stack plot of normalized traces as they are used as the last step in the modified signal time algorithm, for all stations in all events.

For a more clear distinction between the two methods, figure 9 puts the original method up against the pulse start-based one for the representable threshold value of 0.35. Again, the original method clearly results in a smaller angle between both estimates of the shower axis.

5 Conclusion

The results of using the described modification to the standard Offline reconstruction implementation shows that no significant improvement was achieved by switching from a pulse maximum event definition to a pulse start version. Though an improvement could have been expected because the pulse start is a better parameter to base directional reconstruction on. From this, we conclude that there are other noise effects that influence the pulse start more than its maximum. Possibly it's the background noise that has more influence for low signal-to-noise values — such as the pulse start — than for the more clearly defined high SNR values like the pulse maximum.

The results in table 1 indicate that the error in the shower axis is $\sim 12.4^\circ$ using an optimal threshold value. Thus, the uncertainty in the average angle between the axes used in the results for our set of 45 events is $12.4/\sqrt{45} \approx 1.9^\circ$. Since the difference in the average angle we found is $\sim 4.5^\circ$ at best, we can distinguish the original algorithm as the best in the conditions we compared.

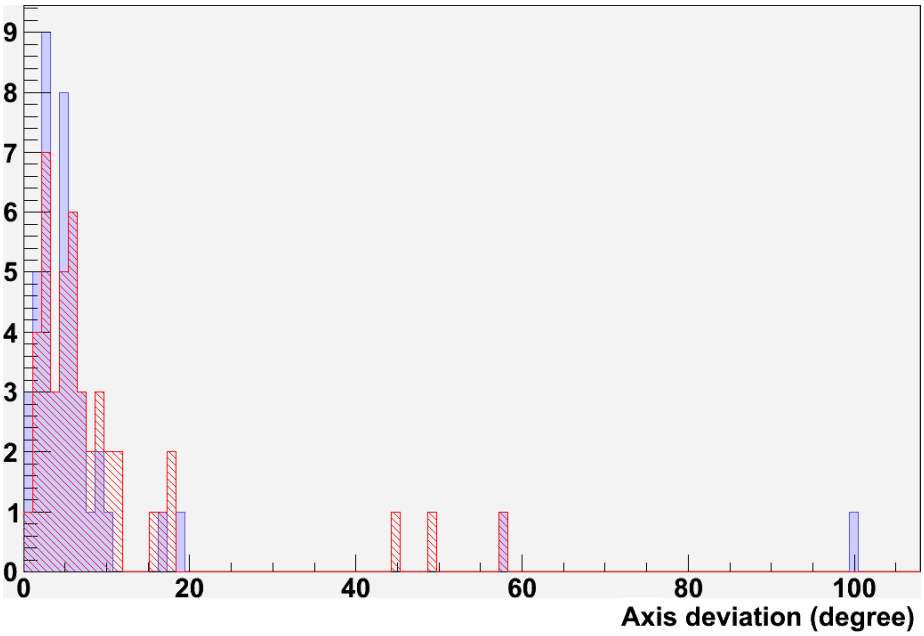

Figure 9: Distribution of angles between shower axes as reconstructed using radio data and hybrid data as a reference, for the original reconstruction (shaded blue) and the modified implementation (striped red) using a threshold value of 0.35.

At any rate, we've shown that for the case of the station signal arrival time evaluation, the current method cannot easily be improved upon. This indicates that the radio reconstruction is maturing as a project. As the results for radio detection at the Pierre Auger Observatory come closer to the established hybrid method that relies on surface Cherenkov and fluorescence detectors, the new radio detection is growing into a valuable addition to the current methods employed by the Pierre Auger collaboration.

References

[1] Carola Dobrigkeit for the Pierre Auger Collaboration. Recent results from the pierre auger observatory, 2009.

[2] James W. Cronin. The highest-energy cosmic rays. 2004.

[3] Diego Harari, for the Pierre Auger Collaboration. Search for correlation of uhecrs and bl lacs in pierre auger observatory data. 2007.

[4] David J. Fegan. Detection of elusive radio and optical emission from cosmic-ray showers in the 1960s. 2010.

[5] John L. Kelley for the Pierre Auger Collaboration. Aera: the auger engineering radio array. 2011.

[6] T. Huege, for the Pierre Auger Collaboration. Radio detection of cosmic rays in the pierre auger observatory. 2009.