University of California

Los Angeles

Small Scale Angular Clustering of the Highest Energy Cosmic Rays in the Surface Detector Data of the Pierre Auger Cosmic Ray Observatory

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Physics

by

Tohru Ohnuki

2005

The dissertation of Tohru Ohnuki is approved.

Jay Hauser

Tatsuo Itoh

David Saltzberg

Katsushi Arisaka, Committee Chair

University of California, Los Angeles

2005

Table of Contents

List of Figures

List of Tables

Acknowledgments

Chapter 4 (The Engineering Array) contain a version of the following internal Auger paper: GAP-2002-059, co-written by T. Ohnuki, D. Barnhill, A. Tripathi, T. Willse, K. Arisaka, D. Warner, J. Horton, titled Photomultiplier Assembly Environmental Protection.

Chapter 6 (Dynode-Anode Ratio) contains a version of the Auger internal paper: GAP-2003-063, co-authored by T. Ohnuki, D. Barnhill, A. Tripathi, K. Arisaka, and titled Determination of Dynode/Anode Ratio from Muon and T3 Data.

Chapter 7 (Simulation of Tank Response) contains a version of the Auger internal paper: GAP-2004-043, co-written by Tohru Ohnuki, Gonzalo Rodriguez-Fernandez, David Barnhill, Arun Tripathi, Tom McCauley, Tom Paul, Katsushi Arisaka, and titled A Systematic Comparison of Three Tank Simulators (G4/SDSim/FastSim) in the DPA Framework.

Chapter 8 (Small Scale Clustering Analysis) includes assistance from Dmitry Semikoz, Weichung Ooi, Antoine Calvez and Steven Flores who worked with us during the summer of 2004. We would like to acknowledge the UCLA Physics and Astronomy Department Research Experience for Undergraduates Summer Program and the National Science Foundation for sponsoring Antoine Calvez and Steven Flores.

Chapter 9 (Auger Clustering Analysis Preliminary Results) contains a version of an internal Auger paper: GAP-2005-026 authored by myself, Dmitri Semikoz, Weichung Ooi, David Barnhill, Antoine Calvez, Steven Flores, Joong Lee, Matt Healy, Arun Tripathi and Katsushi Arisaka, titled Search for Small Scale Angular

Clustering of Cosmic Rays Above 10EeV in the Auger Surface Detector Data.

Vita

1967	Born, Pasadena, California, USA.
1986-1996	Broadcast Engineer, KROQ-FM, Los Angeles
1999	BA in Physics, Occidental College
1999–2001	Teaching Assistant, UCLA
2002	MS in Physics, UCLA
2001–2005	Graduate Student Researcher, UCLA

Publications

Measurement of Carbon Disulfide Anion Diffusion in a TPC. T. Ohnuki, D. P. Snowden-Ifft, C. J. Martoff Nuc. Instr. and Meth. in Phys. Res. A 463 (2001) 142-148 also xxx.lanl.gov physics/0004006

A systematic study of large PMTs for the Pierre Auger observatory. A. K. Tripathi, S. Akhanjee, K. Arisaka, D. Barnhill, C. D'Pasquale, C. Jillings, T. Ohnuki and P. Ranin Nuclear Inst. and Methods in Physics Research, A 497 (2002) 331 - 339

Study of Long Term Stability of the Pierre Auger Surface Detector Using Muon Events. Tohru Ohnuki, Aaron Chou, William Slater, Arun Tripathi, Katsushi Arisaka. Proceedings of the 28th International Cosmic Ray Conference (2003) 817-820.

Abstract of the Dissertation

Small Scale Angular Clustering of the Highest Energy Cosmic Rays in the Surface Detector Data of the Pierre Auger Cosmic Ray Observatory

by

Tohru Ohnuki

Doctor of Philosophy in Physics

University of California, Los Angeles, 2005

Professor Katsushi Arisaka, Chair

***warning, this is cut-n-paste We present preliminary results of small scale angular auto-correlation studies using the first year of Auger data at energies greater than 10 EeV. First, to confirm that our programs computed the correct probability, we reproduced a probability scan over energies and separation angles using the AGASA data. We then performed auto-correlation studies on the Auger data using the same programs. Finally, Monte Carlo simulations were done to determine the statistical penalty for using this type of scan. Our results with one year of Auger data show no significant auto-correlation.

CHAPTER 1

Introduction

1.1 History of Cosmic Rays

1.2 Origin oc Cosmic Rays

CHAPTER 2

Current Experiments and Results

2.1 Two Techniques – Surface Detector and Fluorescence Telescope

2.2 AGASA

2.3 HiRES

2.4 Conflict Between AGASA and HiRES

CHAPTER 3

The Pierre Auger Observatory

3.1 Motivation – History

3.2 Design – A Hybrid Detector

3.2.1 The Surface Detector

3.2.2 The Fluorescence Detector

CHAPTER 4

Photomultiplier Environmental Protection

4.1 Introduction

The surface detector of the Pierre Auger Observatory consists of 1600 individual detectors spaced at 1.5km over a region of the Pampa Amarilla in Argentina. Each detector houses a plastic bag containing purified water which is observed by three PMT's attached to windows on the bag. The PMT's are part of an assembly which includes a base with on-board HV generation, signal amplification and monitoring, the design for which can be found in [2, 3]. The design for the Engineering Array used socketed PMT's and connectors on the base for ease of parts exchange, but unfortunately this reduced reliability. It was decided at the 2001 Surface Detector Electronics workshop in Orsay that the production design would include modifications to improve reliabilty.

4.2 Theory

Obviously, water and electronics don't mix. Actually, the reality is somewhat more subtle. Apart from failure of mechanical components due to water, the main damage mechanism is allowing movement of ions. This current not only affect the circuit in operation, but allows metal etched from the circuit traces to cause breaks in traces or be redeposited leaving a permanent conductive path.

One way to prevent this occurrence is hermetic sealing of the electronics from the atmosphere using metal and ceramic enclosures, but this is not feasible due to cost and difficulty. Plastic enclosures won't work as it turns out that plastics allow water molecules to diffuse through them quickly in comparison to metals and ceramics [4]. So even with perfect sealing of plastic enclosures, water can transpire and damage electronics. However, electronic encapsulating materials protect circuits by preventing liquid water from contacting the board and limiting the movements of ions within the material.

4.3 Encapsulation Methods

There are three main techniques for protecting electronics by encapsulation: Conformal coating, dip encapsulation and potting, in increasing order of material thickness and application difficulty. A conformal coating, as the name suggests, is a thin layer of plastic that conforms to the board and its components. It can be sprayed or painted on and is usually like a paint or varnish with a thickness on the order of tenths of a millimeter. Dip encapsulation involves dipping the assembly and letting the coating flow off leaving a $\sim$1 mm layer. Potting is a casting of the encapsulant allowing infinite thickness. However, this technique requires a containment volume around the assembly and the material be mixed and dispensed in a controlled manner.

Potting was chosen for the preproduction design for the following reasons. Though conformal coatings may offer some protection from water, they have difficulty coating sharp corners with any uniformity, such as on large surface mount capacitors. Also, since the material contracts while drying, this can lead to voids under components where water can collect. Dip encapsulation affords more protection than conformal coatings but in order to leave a thick coating it

must be viscous, again leading to a trapped void problem. Potting solves these issues along with providing physical support for the base and neck of the tube as will be described later. Potting is also the gold standard of protection for experiments, such as Kamiokande, where large PMT's are exposed to water.

4.4 Encapsulation Materials

There are many materials used for encapsulation and they can be broadly categorized by stiffness: Rigid, flexible and gel. Rigid materials such as epoxy, polyester and polyurethane provide excellent protection, mechanical rigidity and are relatively inexpensive. Silicones, available in both flexible and gel consistencies, also provide excellent protection and some mechanical support but are more expensive. Silicones are also removable/repairable whereas the rigid materials are not. This ultimately drove our decision towards silicones.

The preproduction base is fairly complex, having an on-board HV generator and op-amps for monitoring and control. In other experiments where the base was a simple resistor divider with external HV, it was economical to pot the base in a non-repairable material since a failure was most likely due to the PMT and the entire module was replaced. Since the reliability of the Auger base is difficult to determine due to the newness of some of the components and there is some economy in repairing a PMT assembly, a removable material was needed. Gel silicones have the additional advantage that they are self repairing, but were rejected ultimately because they don't provide enough mechanical stability.

The rigidity of the encapsulant was also important from a mechanical standpoint. In order to provide for manufacturing tolerances and ease of installation, a material that had some compliance was preferred over a rigid material. There

Property	Value	Unit
Mix Ratio	20:1	
Viscosity	1500	Centipoise
Specific Gravity	1.00	
Hardness	18	Durometer, Shore A
Temperature Range	-54 to 204	Degrees C
Thermal Conductivity	0.17	W/m deg. K
Coefficient of Thermal Expansion	29×10^{-5}	
Dielectric Strength	15.7	kV/mm
Volume Resistivity	10^{13}	Ω-cm
Work Life	1.6	Hours
Curing Time	24	Hours at 25 deg. C

Table 4.1: RTV-12 silicone properties from [1]

is also less concern for differential thermal expansion rates causing mechanical stress as the estimated expected temperature range of the assembly is -10°C to +50°C.

4.5 Choice of Silicones

The material that was chosen for preproduction is RTV-12 made by General Electric. Silicones from both GE and Wacker of Germany were evaluated, but RTV-12 has many advantages which led to its selection. Its properties are listed in Table 4.1 [1].

The low viscosity and long pot life ensures release of bubbles entrained during mixing and dispensing, eliminating the need for vacuum degassing. It is clear, allowing visual inspection of potting quality and later component inspection for repair. Cure time is short enough such that an overnight cure is sufficient for handling, allowing a one day pot turnaround and the mechanical stiffness is sufficient to protect the cable entry into the material. Since it is a condensation cure silicone, it does not have 'cure inhibition' a problem where the material

fails to cure when in contact with certain common electronics materials. Most importantly, RTV-12 does not require a primer. Many silicones do not adhere well and require a primer which would add a processing step, drying time and need additional toxic materials handling. Because the 20:1 mix ratio is rather high and difficult to mix by hand, a piston mix-meter machine was obtained to ensure proper mixing and dispensing.

4.6 Assembly Design

To allow for the various requirements of shipping, assembly and deployment, groups at UCLA, CSU and Photonis developed a design for the PMT assembly and packing. Since it was agreed at the SDE workshop in Orsay that to increase reliability there would be no connectors within the tank, Photonis proposed providing the PMT's with 'flying leads' which would be soldered to the base board. The board would be supported by a plastic standoff which would be glued to the tube, the leads being potted within the standoff. To provide a enclosed volume for potting and a mechanism for aligning the PMT assembly within the assembly cover (fez) a collar was designed with a flared flange. This collar would be glued to the neck of the PMT while being held in an alignment fixture, ensuring a proper fit. The flange would allow installation of the tube without having to align the PMT on site, the mechanics of the collar and cover providing the fixation (no more PAF[1]). For the initial ∼140 assemblies the collar was made of a blue polystyrene foam, the production collars are injection molded ABS. In addition, the packing foam was designed in such a way as to provide support of the PMT while being shipped from Photonis to UCLA and also to Malargue

[1]PMT Assembly Fixture

after the tube was tested and assembled with collar etc.

4.7 Environmental Testing

To ensure that the design and materials chosen for the PMT assembly would withstand the conditions in a tank in the field, a environmental test chamber was constructed. It is capable of a temperature range of -15°C to +80°C at a slew rate of ∼1.5°C/min. It is controlled by an industrial ramping temperature controller which allows one to set the temperature profile and dwell times.

As a check that the initial design would not fail, a tube with nominal properties was selected from the production stream and tested with the chamber under conditions suggested in [5]. This tube was subjected to multiple cycles from -10°C to +65°C with 30 minute dwells at each extreme and with a ramp of 1°C/min. A burn-in period at +65 for 16 hours was also performed. The tube survived without significant changes in performance. It was planned to pull a tube from each batch and perform a similar test, but this was not done primarily since the initial design with the blue styrofoam collar is thermally very different from the final black ABS collar design. Tests will be performed with the new collar design as quantities become available.

4.8 Conclusion

In 2002, it was decided that in order to facilitate tank deployment and reduce PMT processing costs the testing and potting of PMT would be done on site in Argentina. The testing system and potting equipment were shipped and in early 2003 production was started.

To date, approximately half of the PMTs for the array have been tested and deployed. Of these, less than 50 have been returned from the field with problems, though none have been identified as having failed due to environmental protection failure.

CHAPTER 5

Muon Spectrum Calibration

5.1 Calibration of the Auger Surface Detector

5.2 Determining the Muon Hump

5.3 Time Trend of the Muon Hump in the Engineering Array

5.4 Monitoring of the Array Through Muon Data

5.5 Epilogue

CHAPTER 6

Dynode-Anode Ratio

6.1 Introduction

The energy calibration of the surface detector of the Pierre Auger Observatory is made considerably easier by the ubiquitous cosmic ray muons. The muon angular distribution in conjunction with the water tank geometry provide an unambiguous signal with which to calibrate the detectors. The method of analysis involving the tank muon spectrum has been described elsewhere [6]. However, this muon data is taken on the high gain dynode channel whereas large shower physics data will be taken on the low gain anode channel [1]. Because of this difference, the dynode/anode gain ratio is an important contribution to the uncertainty of the shower energy.

The dynode/anode ratio is determined by several factors: individual tube characteristics, high voltage applied to the tube, the design of the base dynode resistor chain and the amplifiers and filters between the base and the FADC. Of these contributions, the individual tube behavior and high voltage applied affect the dynode/anode ratio most significantly and must be determined for each tube.

The overall gain of the tube is typically given by: $Gain = HV^{Cn}$ where HV

[1] These two channels are digitized by 10-bit FADC's with a nominal gain difference of ∼32 (5-bits) between them.

is total applied high voltage, C is a constant (for the Photonis tube used in the Auger array typically about 0.5) and n is the number of dynodes. The gain between each dynode is given by an expression of the form: $\delta \sim \Delta V^C$ where ΔV is the voltage between the dynodes. The gain between the last dynode and anode is given by:

$$\frac{D}{A} = \frac{(\delta - 1)}{\delta} \tag{6.1}$$

with typically $\delta \sim 5$. Clearly, the dynode/anode ratio is weakly dependent on the high voltage. For the Auger tube base, this ratio is then multiplied by the dynode amplifier gain (nominally 40) giving an overall ratio of ∼32. Further details regarding tube specifications and testing can be found in [7].

6.2 Muon Data Analysis

Data from the raw muon traces recorded on each tank in the engineering array during the period from May 2002 to June 2003 were taken under the following conditions:

- To initially set high voltages and trigger levels the rate based technique described in [8] was used.
- During operation, the onboard computer adjusts threshold and provides calibration constants continuously [8].
- 1000 events are recorded every four hours over a period of a few seconds.
- The threshold was lowered to 0.15 VEM[2] on each PMT and threefold co-

[2]A convenient unit for calibration and trigger setting is the *Vertical Equivalent Muon*, the amount of energy deposited by a vertical muon traveling the 1.2 meter height of the water in the tank, about 240 MeV.

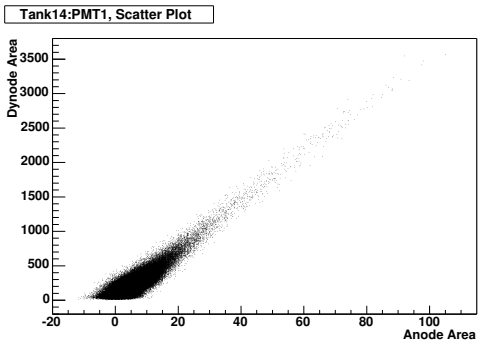


Figure 6.1: Scatter plot of Dynode/Anode Ratio, 1 tube for 1 month of data

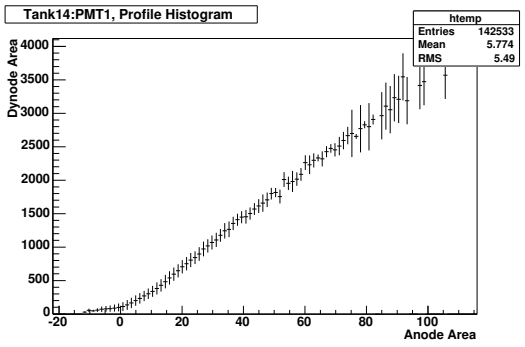


Figure 6.2: Profile histogram of the data in Figure 6.1

incidence was required.

The pedestal or baseline is the quiescent value that the ADC records when there is no signal present. This is computed first as subsequent analysis needs to have the baseline subtracted for charge integration. The pedestal was determined by taking the first sample bin from each event and averaging for all events in a time period. This technique gave pedestal values within 0.02% to the ones given by fitting the histogram of the first few bins to a Gaussian as described in [9].

Each event then has both dynode and anode traces plotted after pedestal subtraction and the area under the curve integrated. For a given time period, these data are plotted yielding a scatter plot (Figure 6.1) This can be fitted to linear function to determine the dynode/anode ratio. Only events with a peak dynode value of less than 900 were used to eliminate dynode saturated events. To ensure that the fitting was performed on the linear portion of the plot, the fit was constrained to be greater than 20.

6.3 Issues with the Analysis

Originally, profile histograms were used giving plots like Figure 6.2. Fitting a linear function to the profile histogram turned out to be difficult due to two issues: The non-zero values at the origin, and how profile histogram determines errors. The skew of points at the origin on the profile plot is due to the asymmetric 'bulb' at the origin on the scatter plot. This is caused by the much larger uncertainty in the integral on small signals on the anode channel. If the least count is one for both channels, the possible integrated error is $\sim \sqrt{25}$ which is significant for typical anode areas of 50, but not for a typical dynode area of $\sim$2000. This problem was somewhat resolved by requiring a minimum anode value.

The second issue regarding the fits to profile histograms was more subtle. In Root (we used version 3.05 for this analysis), generally the fitter will weight points according to the error of the point. In the case of profile histograms, Root would assign an average error value for bins that contained only one point or a few points with the same value. The linear fit would then be skewed by this uneven weighting. It was found that if the histogram was constructed with finer binning, the fit improved since the statistics per bin would go down and more of the bins would be evenly weighted. This problem was resolved by fitting to a scatter plot where each point would be equally weighted.

However, there still remained an issue regarding the cuts applied to fitting the dynode/anode scatter plot in addition to the anti-saturation cut (peak on dynode < 900). A study was performed on the effect of constraining the anode data to be greater than some value. As might be expected, the dynode/anode slope changes drastically if the events dominated by noise on the anode (value $<$ 15) are included. As the cut is made at larger and larger anode values, the slope decreases. Figure 6.3 (red line) show examples of these effects (cuts greater than

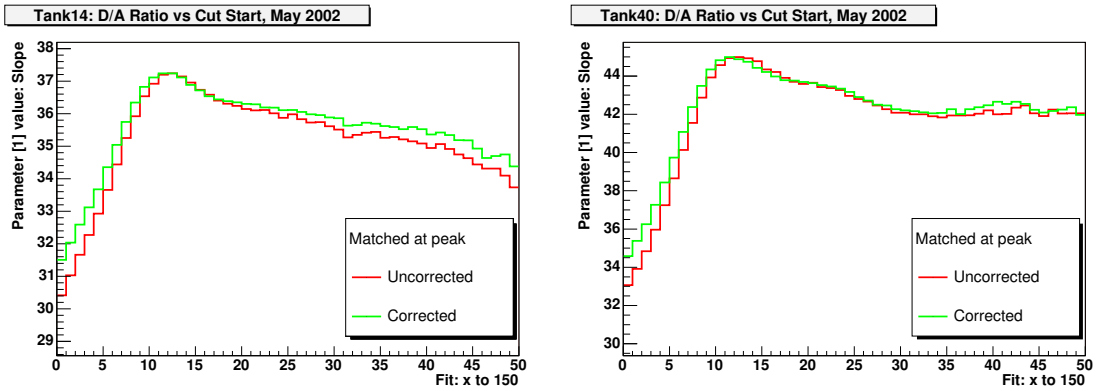


Figure 6.3: Examples of Dynode/Anode Ratio as a function of minimum anode signal for two tanks, 1 month of data

50 were not explored as statistics become poor). Since it has been observed that the EA electronics have an undershoot that is proportional to the signal size, it was thought that this droop in the slope could be due to underestimating the integrated pulse area. A correction factor was added from a linear function fit to the pedestal change versus pulse size described in [10]. But as can be seen in Figure 6.3, this correction (green line) did not solve the problem, suggesting that the problem is elsewhere in the hardware. If the amplifiers and filters before the ADC were linear, one would expect that the slope value would stabilize after the influence of the noise was cut. However, this droop in the slope value suggests some nonlinearity in the electronics. This nonlinearity can also be seen in the profile histogram of Figure 6.2. For the results reported below, a minimum anode area of 20 was used.

Finally, there is the issue of the functional form of the dynode/anode ratio. Ideally, it would be a linear function which goes through the origin. Both fits where the y-intercept was set to zero (constrained) and allowed to float (unconstrained) were studied with results in Figures 6.4 and 6.5. In the unconstrained fit, the intercept varied without settling to any value and the dynode/anode ratio decreased for increasing anode values (Fig. 6.6, left). In the constrained fit, one

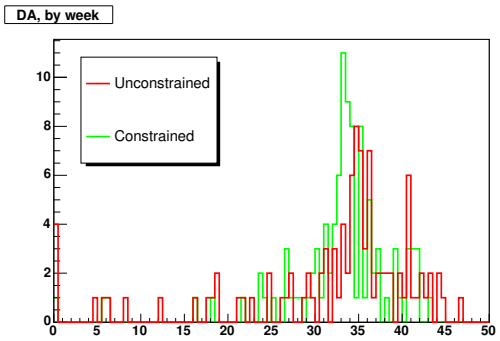


Figure 6.4: Histograms of constrained and unconstrained fit, all tanks, all tubes.

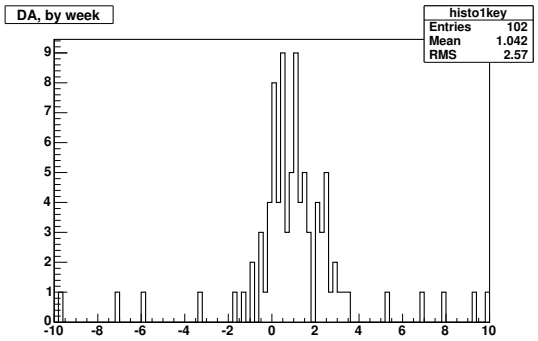


Figure 6.5: Histogram of the difference between constrained and unconstrained Dynode/Anode Ratios, all tanks, tubes.

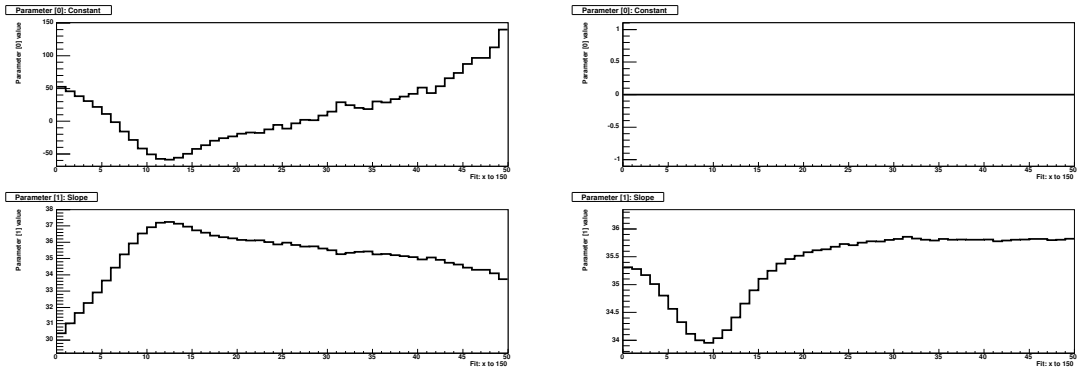


Figure 6.6: Effect of anode cut on linear fit with floating (left) and fixed (right constant (X axis is minimum anode signal)

can see that the slope value has a smaller variation with an asymptotic value as the cut bin is increased (Fig. 6.6, right). This is to be expected as increasing the minimum anode area would restrict the fit to fewer higher value data and the origin. This is the kind of behavior expected if the pedestal subtraction works correctly and the amplifiers were linear. In real life, we cannot constrain the intercept to zero or assume linearity. The difference between the two gives an idea of the systematic error.

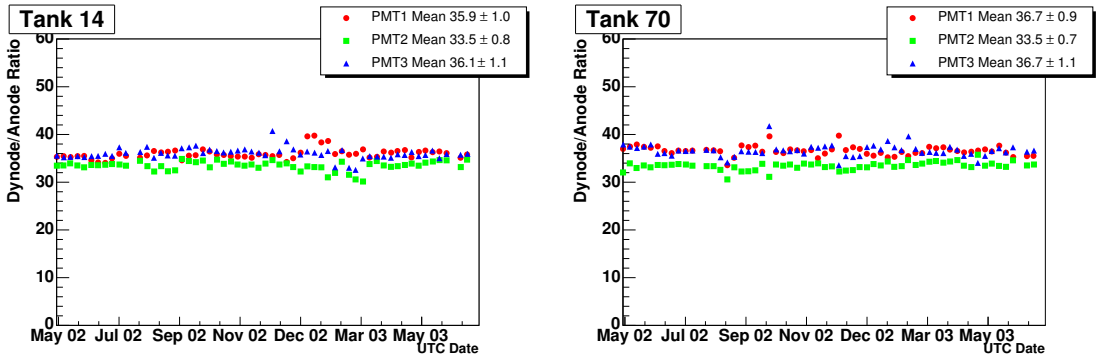


Figure 6.7: Time trend of D/A ratio over one year for two selected tanks

6.4 Results

For muon data, taking data in one week intervals gave good results: ∼40 files of 1000 events. For well behaved tanks (ones with consistent muon data) the scatter of the dynode/anode ratio was on the order of 2–3% over the period May 2002 to June 2003 (Figure 6.7). Problem tanks showed a much larger scatter, in some cases up to 10%. The gain change made to the array in July 2002 is not clearly visible in these plots which is expected since the gain change of five times should only change the dynode/anode ratio by about a percent.

T3 trigger physics data did not give good results even when taking data in one month intervals. For an average tank, there were less than 1000 events per month. In addition, the type of events that passed the criteria of not saturating the dynode were mostly muons and had a maximum of only about twice the signal as from muon data. Thus the spread in values using this data is significantly worse as can be seen in Figures 6.8 and 6.9 (this is consistent with results in [11]).

Comparing the muon data dynode/anode ratio with local station computed values show them to be mostly consistent with some non-Gaussian tails (Fig. 6.10). However, the scatter in the local station value is smaller due to possibly

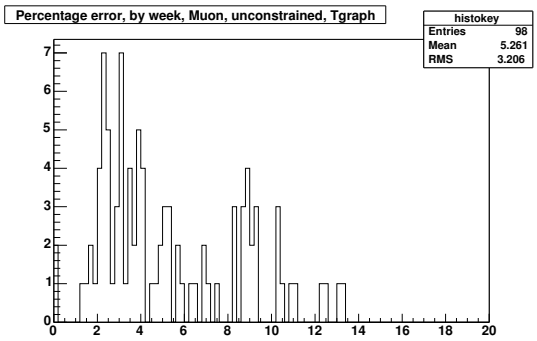


Figure 6.8: Histogram of D/A ratio scatter of muon data in percent, all tanks, tubes

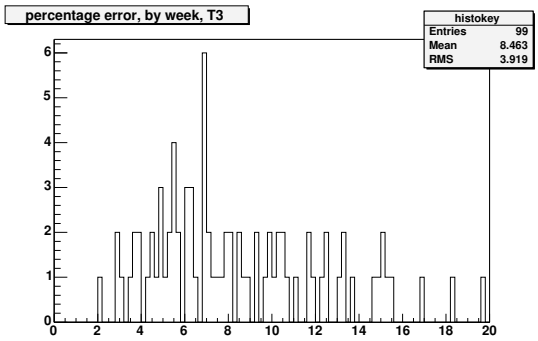


Figure 6.9: Histogram of D/A ratio scatter of T3 data in percent, all tanks, tubes

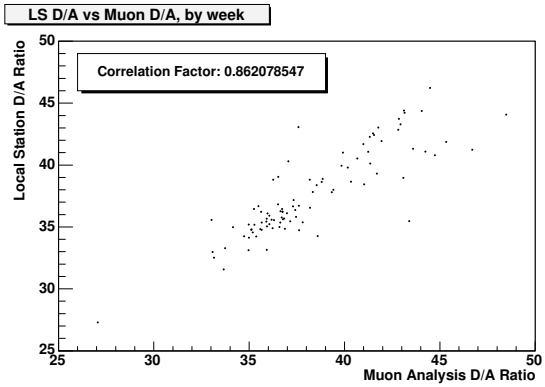


Figure 6.10: Correlation between local station and muon analysis D/A ratio, all tanks, tubes.

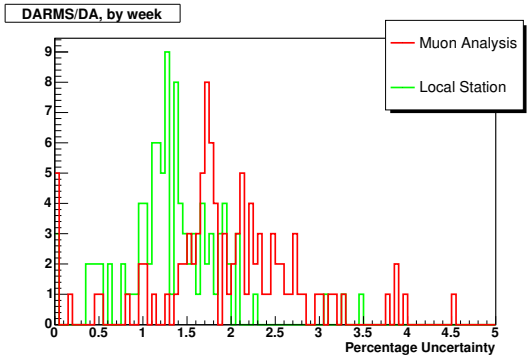


Figure 6.11: Histograms of D/A ratio RMS by muon analysis and local station, all tanks, tubes.

different technique or selection of larger events (Fig. 6.11).

6.5 Results Summary:

- Using an offline analysis allows us to determine the ratio independently for comparison with local station values but with a scatter of up to 10
- Neither the current muon trigger or the T3 data supplies data that would allow better determination of the dynode/anode ratio.

- There is some non-linearity that makes determination of the ratio dependent on where cuts are placed in the data.
- The analysis shows that the dynode/anode ratio is still not understood well enough to give reliable and accurate results.

6.6 Discussion

Since we do not yet have muon data from the pre-production array, this analysis is only applicable to the engineering array. However, the engineering array is becoming less important as more and more pre-production tanks become operational. We must now look at how to determine the dynode/anode ratio on the pre-production tanks which have a new design that has different phototubes, base and front-end electronics. Unfortunately, the slow buffer muon data and the LED data currently available cannot be used for precision measurement of dynode/anode ratio.

First, the slow buffers on the new unified board only store the dynode channels, so this data, while good for muon hump measurement, cannot be used for the ratio.

Second, the LED flasher data currently being used for dynode/anode ratio in pre-production is not optimal. The design of the LED flasher is such that the short light pulse (tens of nanoseconds) simulates a muon pulse. This makes it impossible to get larger integrated signals before dynode saturation. Analysis from the preproduction tanks monitoring data shows a systematically higher ratio than that measured when the phototube assemblies were tested [9]. So for pre-production we would propose the following:

- A special dynode/anode ratio trigger with a high trigger threshold set at 2

to 5 VEM and requests data from all stations. This should provide small shower events that have a larger EM component, thus pulses that have larger area for a given peak value. Since the trigger rate will be much lower than with the muon trigger and require more array dead time, this data-taking needs to be done now during pre-commissioning of the array. Results of this analysis can be folded back into on-tank software.

- Muon data-taking, with the same trigger conditions to the engineering array to cross check that the special ratio trigger described above is working correctly.

- Modification of the LED flasher electronics to provide a longer pulse to increase data at large signal values before dynode saturation. It may be that non-linearity is dependent on pulse width, therefore a study with longer pulses (up to a few microseconds) is necessary.

- Since our result suggests that there is some non-linearity in the electronics, a check of anode linearity at large signals is required. Using the two LED's on the tank, the linearity can be measured with the same A, B, A+B technique used in the PMT testing [7]. Again, dependence on pulse width needs to be tested. In conclusion, the goal of the Auger experiment is to measure accurately the shower energy and spectrum. The current scatter in the dynode/anode ratio is larger than our goal to determine detector uncertainties to better than 5%. Every effort must be made to reduce detector uncertainties in the pre-production array before real physics data-taking commences.

CHAPTER 7

Simulation of Tank Response: Comparison of Three Tank Simulations

7.1 Introduction

The Pierre Auger Ground array consists of 1,600 water tanks. These water tanks respond to different particle kinds in different ways. For the electromagnetic component of a shower, the tank behaves as an absorption calorimeter, whereas for a muon, it dramatically enhances the signal due to its long path length in the water. To be concise, a water tank is as a peculiar mixture of an EM calorimeter and a muon detector. To the end, the standard observables such as Lateral Distribution Function (LDF) and S(1000) are not universal, but very detector specific.

This arrangement has important consequences for calibrating the absolute energy. The calibration is basically performed by the so-called Vertical Equivalent Muon (VEM), but information of primary energy of cosmic rays is carried by the electro-magnetic component. Unfortunately, these two signals have very different pulse height and width, and measured by different electronics channels, which are summarized in Table 7.1 below.

For the ultimate goal of energy calibration, all these factors need to be simulated and understood. Therefore tank simulations play a critical role in ensuring

	VEM	Extensive Air Showers
Composition	Muon	EM Shower + Muon
Typical Energy	>1 GeV	1 - 10 MeV (γ, $e^{\pm}$), >1 GeV ($\mu^{\pm}$)
Direction	Vertical	All Angles
Pulse Width	$\sim$65 nsec	$\sim$2 μsec
Pulse Height	1 VEM	$\sim$5000 VEM at S(1000) for 10^{21}eV
Readout	By Dynode	By Anode

Table 7.1: Comparison between VEM and extensive air showers. All these differences must be taken into account for absolute energy calibration

the accuracy of the energy measurement by the ground array. In the past, simulations had been done by various groups1 but comparison of results was difficult, because there was no easy way to share the common input parameters such as:

- Tank Geometry
- Water Absorption Length, Tyvek reflectivity (vs. Wave Length)
- PMT Shape/Diameter, Quantum Efficiency, Collection Efficiency.

Thanks to the newly-developed DPA framework, we are now allowed to compare the publicly available simulation modules using the same input conditions, only changing the module under test. we have therefore decided to compare three modules: Geant4 based G4SIM [12], SDSIM [13] and FASTSIM. In this paper we describe how this was done and report the results. This work should not be construed as a criticism of any of the simulation modules or their authors. Our goal was to cross check and gain confidence that the software is doing what we expect and to find and correct bugs. We were in contact with the authors of the modules and most problems have already been corrected.

7.2 Overview of DPA Based Detector Simulation

7.2.1 The DPA Software

The DPA software package is a framework for analysis and simulation written by the DPA group for use by the Auger collaboration. The framework provides an event structure and methods to pass this structure through modules that perform calculations upon the data in the event. To do complex analyses, a chain of simple modules is created where each module does a portion of the calculation. This compartmentalization allows one to easily swap out modules for comparison and isolate modules for troubleshooting. For this work, a particle generator module was followed by the tank simulation module under investigation and output read by an event dump module. (Fig. 7.1) An example module sequence file follows:

```
<sequenceFile>
  <moduleControl>
    <loop numTimes="1000">
      <module> ParticleInjector </module>
 <!-- <module> FastTankSimulator </module> -->
      <module> G4TankSimulator </module>
 <!-- <module> SDSimTankSimulator </module> -->
      <module> EventDump </module>
    </loop>
  </moduleControl>
</sequenceFile>
```

In this example, the Geant4 simulation is selected, the others are not used.

The particle type, energy, position and direction were specified in another configuration file which is read by the injector. When the simulation is started, an

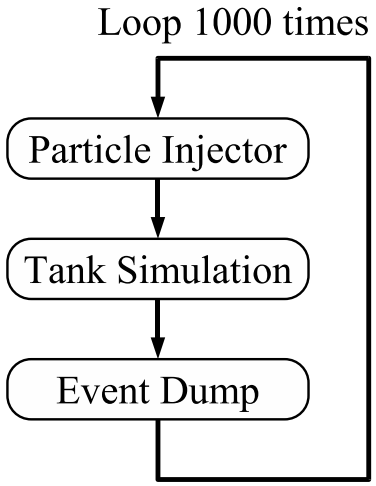


Figure 7.1: DPA Simulation loop

Event is created by the framework which is handed to the first module. The particle injector then either Gets or Makes structures within the Event for injecting the simulated particle (Fig. 7.2).

Next, the tank simulation does a similar procedure to get the particles. After getting the particle's type, energy, position and direction the simulation is run. Geant4 traces the generation and propagation of photons within the tank until they contact a phototube or are lost, simpler simulations generate a list of photons using probability. The simulations then create a time distribution object within the event structure for the photoelectron trace and fill it with the simulated data. After the tank simulation, the framework again hands the Event to the dump modules which drills down into the Event structure to read out the particle properties (as a check) and the photoelectron trace, see Figure 7.3. The trace is histogrammed and its integral and other computed values written to a root file.

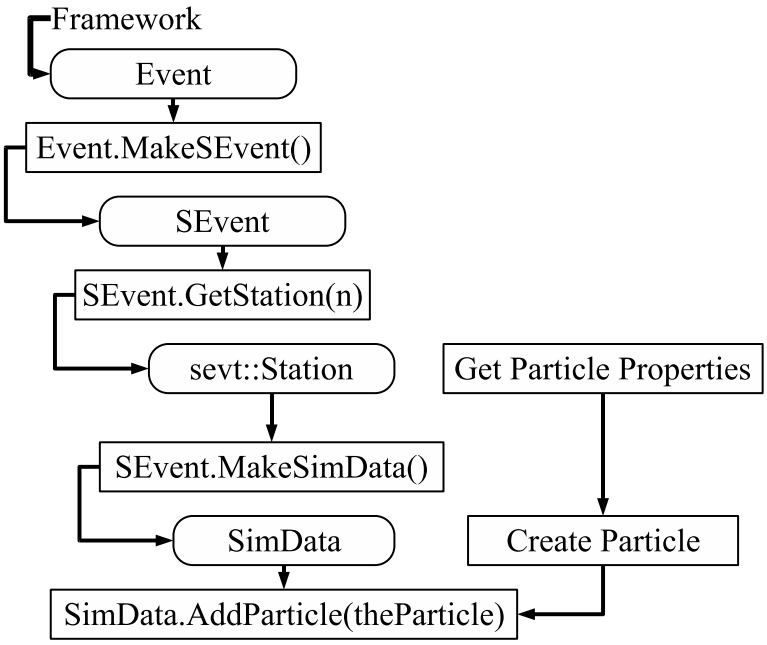


Figure 7.2: Injector Module, receives event from the Framework and injects particle. The square boxes are functions that return the data object in the round boxes

7.2.2 The Modules

The three simulations that were compared were the G4, SDSim, and FastSim. The G4 and SDSim were originally stand alone software but have been ported to the framework. The following comments are based on our investigation of the source code, questions about the details of each module should be addressed to the module author. The G4 simulation uses Geant4 for physics and includes many processes. Geant4 is a toolkit for the simulation of the passage of particles through matter. Geant4 lets us define the detector geometry, the materials and theur physical properties, for instance, optical photon wavelength dependence of water absorption length (G4TankConstruction), energy cuts, turn on/off various process such as muon decay, delta ray production etc. Geant4 follows the track of all particles, with G4Step and G4Track, we can get or modify all the particle

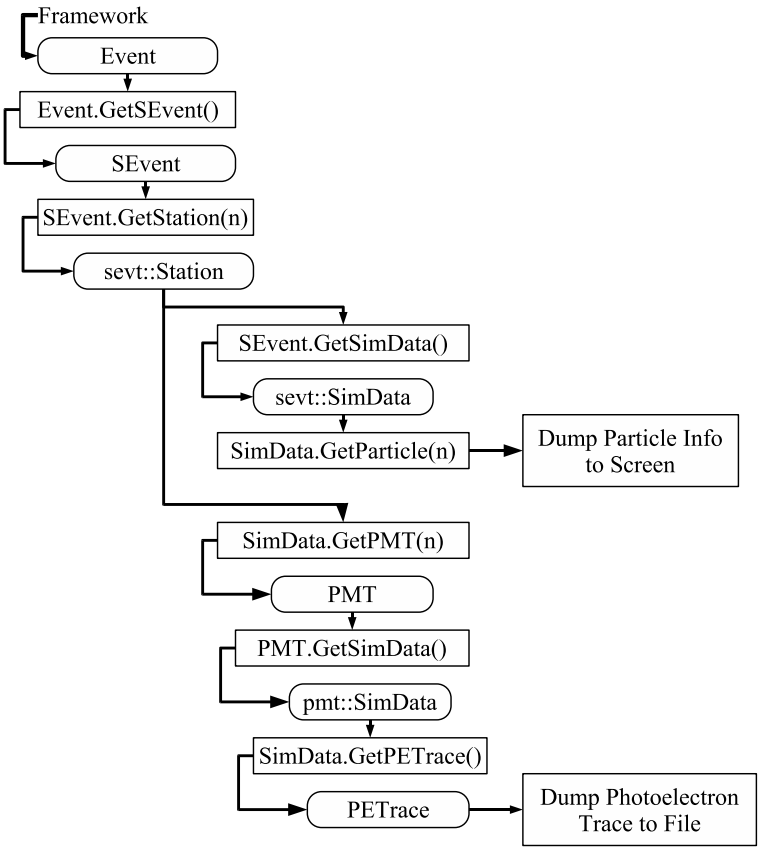


Figure 7.3: Event Dump module, receives event from Framework and dumps data of interest

properties at each step, e.g. kill the particle, change its direction, get its energy and so on. Every trajectory is divided in "steps", this "step length" depends on the material by default, but the user can change this "step size" manually to speed up the simulation. For instance, the class G4TrackinAction is called before and after one particle is created or killed, and the class G4SteppingAction is called at each step of the particle.

With Geant4 we can define the "Sensitivity volume", which means when some type of particle hits this volume, G4 can call a class defined by the user to take some action on this particle. For instance, the photomultipliers, are de-

G4Sim	SDSim	FastSim
750s	20s	3.5s

Table 7.2: Comparison of simulation speed

fine as a "Sensitivity volume", when an optical photon hits this volume then G4TankPMTAction is called to see if this optical photon is detected or not.

For checking that the simulation is behaving correctly, we can interactively from console define all parameters of the simulation, like initial particle position, direction and energy, physical process on/off, etc. For more information:

`http://wwwasd.web.cern.ch/wwwasd/geant4/geant4.html`

SDSim is a simplified version of G4, without all physics process and materials properties and unlike G4 does not raytrace the Cerenkov photons, thus is significantly faster than the G4 simulation. For more information:

`http://lpnhe-auger.in2p3.fr/Sylvie/WWW/AUGER/DPA/`

FastSim is the simplest simulation with only the main physics processes included and is very fast. Due to these differences, the simulations run with the times in seconds shown in Table 7.2 for 100 vertical muons on a 1GHz Pentium 3 class machine.

Note that these times are for 100 particles injected simultaneously, since for individual particle injection, the setup and teardown of the framework is the majority of the time used for the simpler simulations.

7.2.3 Input Parameters

Ideally, all three simulations would get their input parameters from the same source, thus eliminating problems and questions regarding them. However, as

FastSim	SDSim	G4Sim
PhotoElectronRate 80	AbsorptionLengthMax 30	AbsorptionLengthMax 30
	MaxReflectivity = 0.975	MaxReflectivity = 0.98
	CE = 1.3 (NB: this is an unphysical value, used to normalize response after setting above parameters to match G4 values	CE = 0.65
	Relative distributions were added for absorption length and TyVek reflectivity	G4TankConstruction was modified to only retrieve tank geometry from the Offline Database

Table 7.3: Input parameters and alterations to the three simulations

currently implemented, the simulations have different sources and values. All three simulators were modified to set the seed of the random number generator to the system clock to give a realistic random behavior in the random processes. In addition the following values were used in the configuration files, Table 7.3:

These parameters and code changes were made so that a vertical 1 GeV muon through the center of the tank would produce ∼100 photoelectrons per phototube or a total of 300, a number measured on real tanks [14, 15]. In SDSim and FastSim the path length for an incident particle is analytically calculated from the injection point of the particle, so particle injectors were defined to inject at the surface of the tank. This is slightly larger than the water volume dimensions and was chosen for consistency with the Geant4 simulation which includes a ∼2 cm tank wall thickness, see Table 7.4.

Injected particle energies ranged from zero to 4 GeV for all species with more low energy runs for electron and gamma. Though high energy electrons and

	Position (x, y, z)	Direction (x, y, z)
Vertical Injection	0, 0, 1.23 m	0, 0, -1
Horizontal Injection	1.83 m, 0, 0.6 m	-1, 0, 0

Table 7.4: Particle injection positions and directions

Electrons, Gammas (MeV)	Muons (MeV)
0.0, 0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0, 1.5, 2, 3, 4, 5, 7, 10, 20, 30, 50, 70, 100, 200, 500, 1000, 2000, 4000	0.0, 0.1, 0.5, 1, 5, 10, 20, 50, 65, 85, 100, 200, 300, 400, 500, 1000, 2000, 4000

Table 7.5: Energies used for three particle types

gammas are not expected on the ground from showers, these points were taken for completeness. List of energies used, Table 7.5:

7.3 Comparison of Three Simulation Modules

7.3.1 Simulation Data and Fluctuation Excess

Three particle species, electron, gamma and muon were injected at various energies vertically through the center of the tank. The tank's output in photoelectrons was histogrammed and the mean, rms, fitted mean and sigma were output to file (see Fig 7.4). Since the fit was only good for a range of energies where the photoelectron distribution could be approximated by a Gaussian, the fit quantities were not used. The *fluctuation excess* is a measure of scatter in the distribution that is greater than the expected fluctuation due to statistics. This was computed by the fraction $\frac{RMS}{\sqrt{mean}}$. This is somewhat similar to the excess noise factor present in photomultipliers. In the case of water tank response, this excess is due to physics processes such as:

- Detector Geometry - For particle flux at an angle, there are differing path

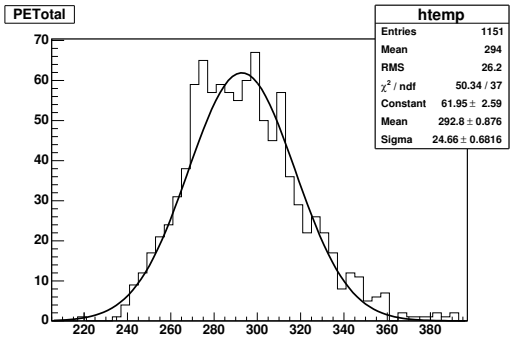


Figure 7.4: Histogram of 1000 1GeV muons vertically through tank in Geant4 simulation. The fit was ultimately not used.

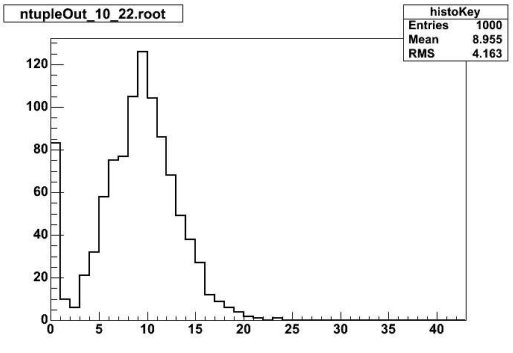


Figure 7.5: Histogram of vertical 10 MeV gamma events in G4, note ∼8% of entries are in the zero bin.

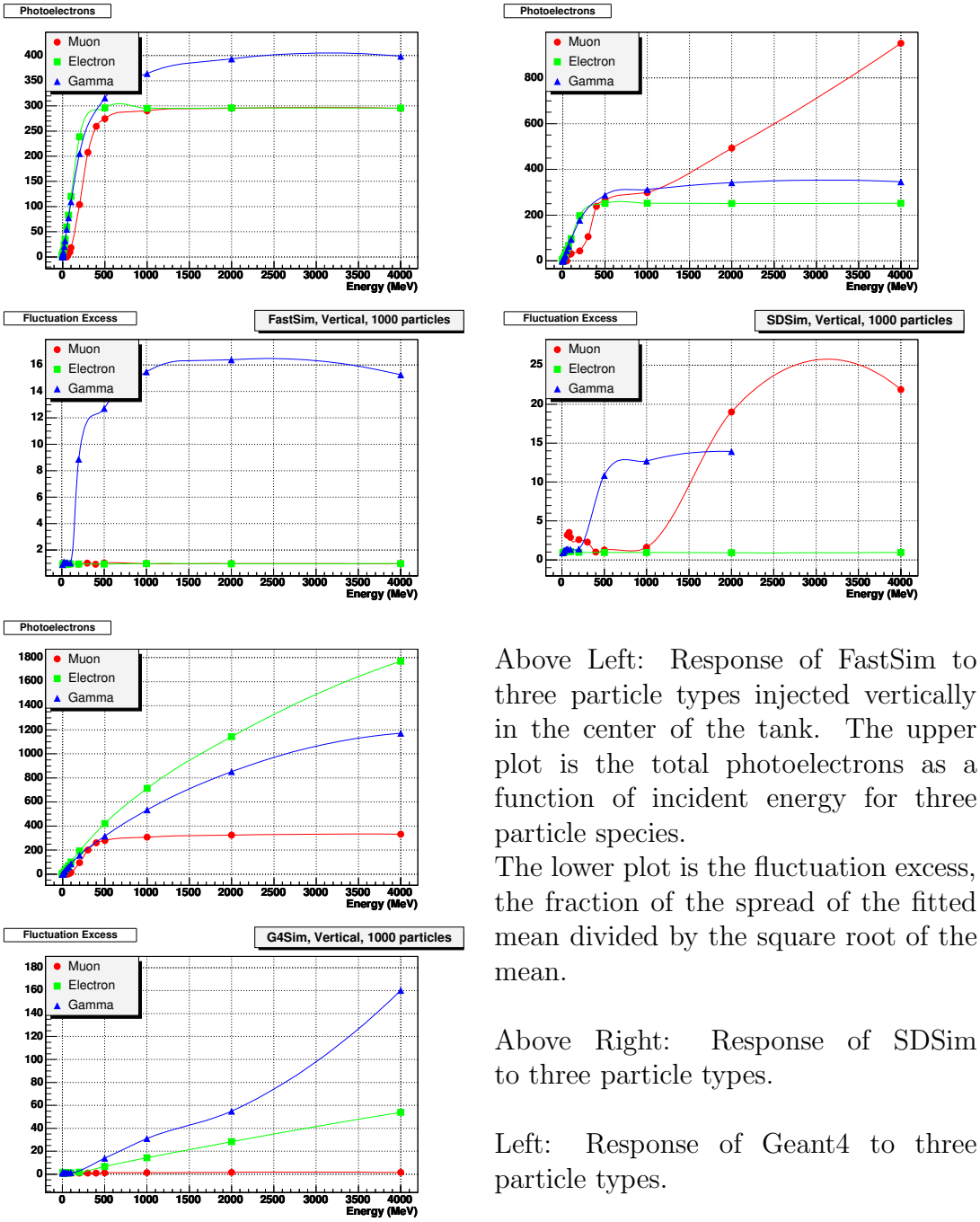
lengths but the average response and fluctuation of this response can be characterized, see Reference [16].

- Multiple Coulomb Scattering
- Incomplete Containment of EM Showers
- Gamma Conversion Probability - If the photon does not interact within the tank, no photoelectrons are detected, see Figure 7.5.

These contribute to the fluctuation in addition to the Poisson statistics due to the number of photoelectrons.

7.3.2 Single Particle Response for Species and Energy

The following plots in Figure 7.6 show the response of each simulation to the injected particles. Note that both FastSim and SDSim seem to treat electrons and gammas like minimum ionizing particles, reaching some plateau in response. Also FastSim seems to have essentially Poisson treatment of electrons and muons shown by the fluctuation excess of ∼1, except for gammas.



Above Left: Response of FastSim to three particle types injected vertically in the center of the tank. The upper plot is the total photoelectrons as a function of incident energy for three particle species.
The lower plot is the fluctuation excess, the fraction of the spread of the fitted mean divided by the square root of the mean.

Above Right: Response of SDSim to three particle types.

Left: Response of Geant4 to three particle types.

Figure 7.6: Response of the three tank simulations to three particle species, mouns, electrons and gammas injected vertically, as linear plots.

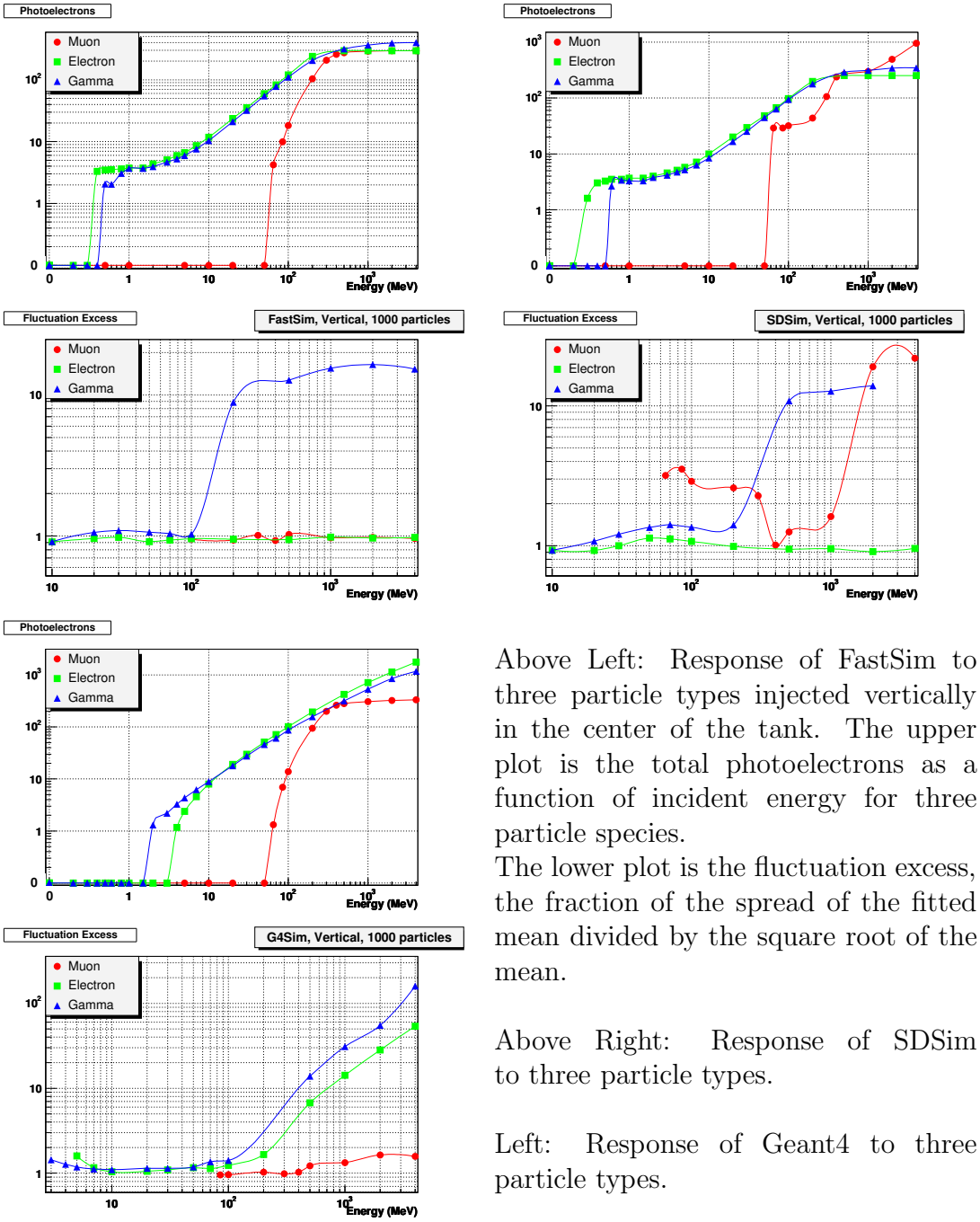
Figure 7.7 are the same data as in 7.6 presented as logarithmic plots which magnify differences in the lower energy range, showing that both FastSim and SDSim have a much lower threshold energy for electrons and gammas than the Geant4 simulation. This is due to the inclusion in Geant4 of the effect of the dead material of the tank. The kink and the rise at high energy in the SDSim muon response is unexplained and has been reported to the author of the module.

7.3.3 Horizontal Response

Again, three particle types were injected horizontally at a height of 0.6 meters along the x axis into the tank. Since the water tank is three times as wide as it is tall, the muon path length and consequently the tank response to muons should be three times larger or about 900 photoelectrons. Only Geant4 shows the correct increase in response with increasing energy, Figure 7.8.

7.3.4 Horizontal Correlation

For horizontal events, the correlation of PMTs was plotted, Figure 7.9. It is interesting that an asymmetry due to the direct light is seen in the PMT response in both G4 and SDSim. This is consistent with work done by the Santiago-Compostela group [17]. The different responses between G4 and SDSim can be explained by the configuration of PMT positions in each simulation. In G4Sim, the tubes are numbered clockwise 1 to 3 whereas in SDSim they were numbered counterclockwise.

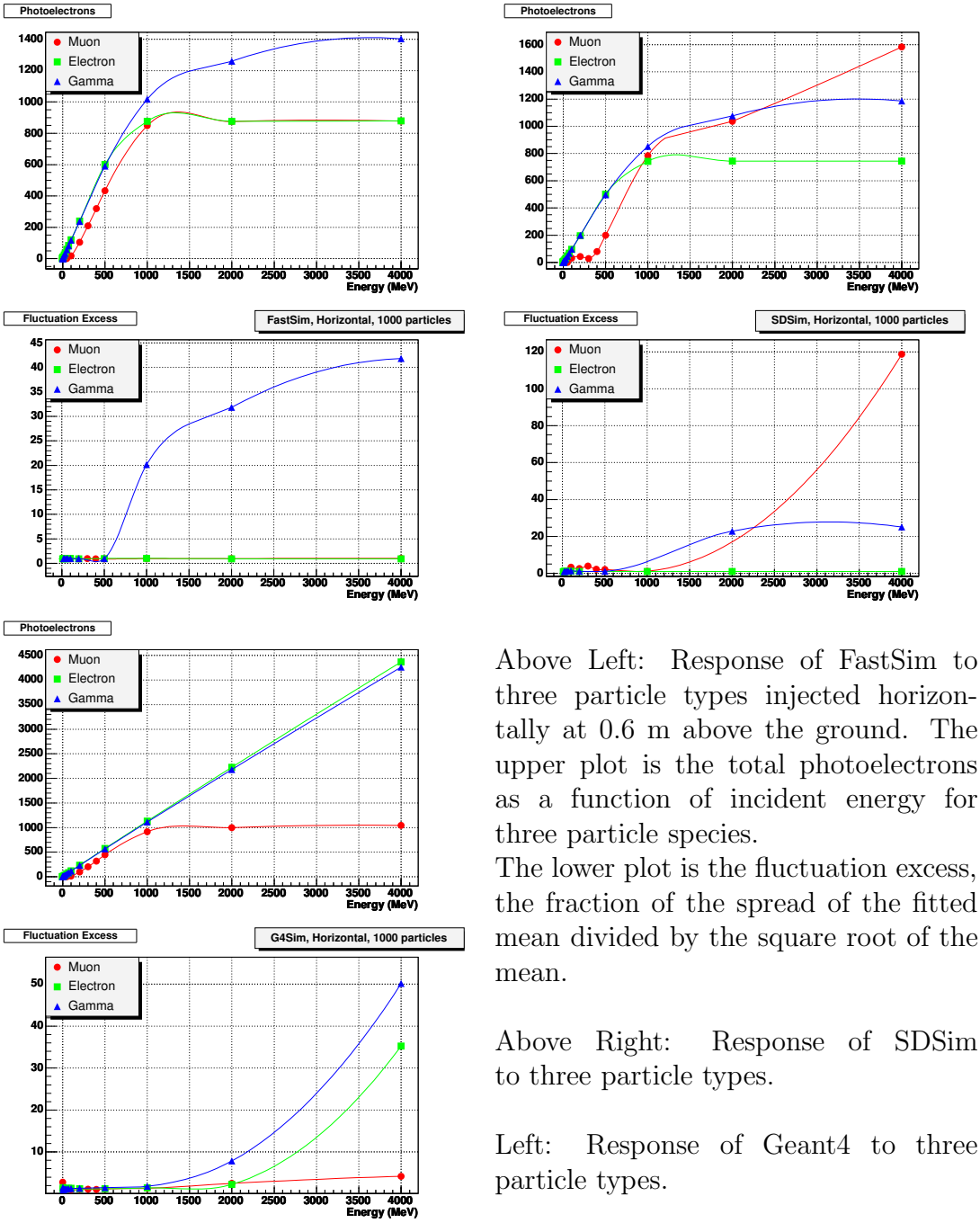


Above Left: Response of FastSim to three particle types injected vertically in the center of the tank. The upper plot is the total photoelectrons as a function of incident energy for three particle species.
The lower plot is the fluctuation excess, the fraction of the spread of the fitted mean divided by the square root of the mean.

Above Right: Response of SDSim to three particle types.

Left: Response of Geant4 to three particle types.

Figure 7.7: Response of the three tank simulations to three particle species, mouns, electrons and gammas injected vertically, as log plots.

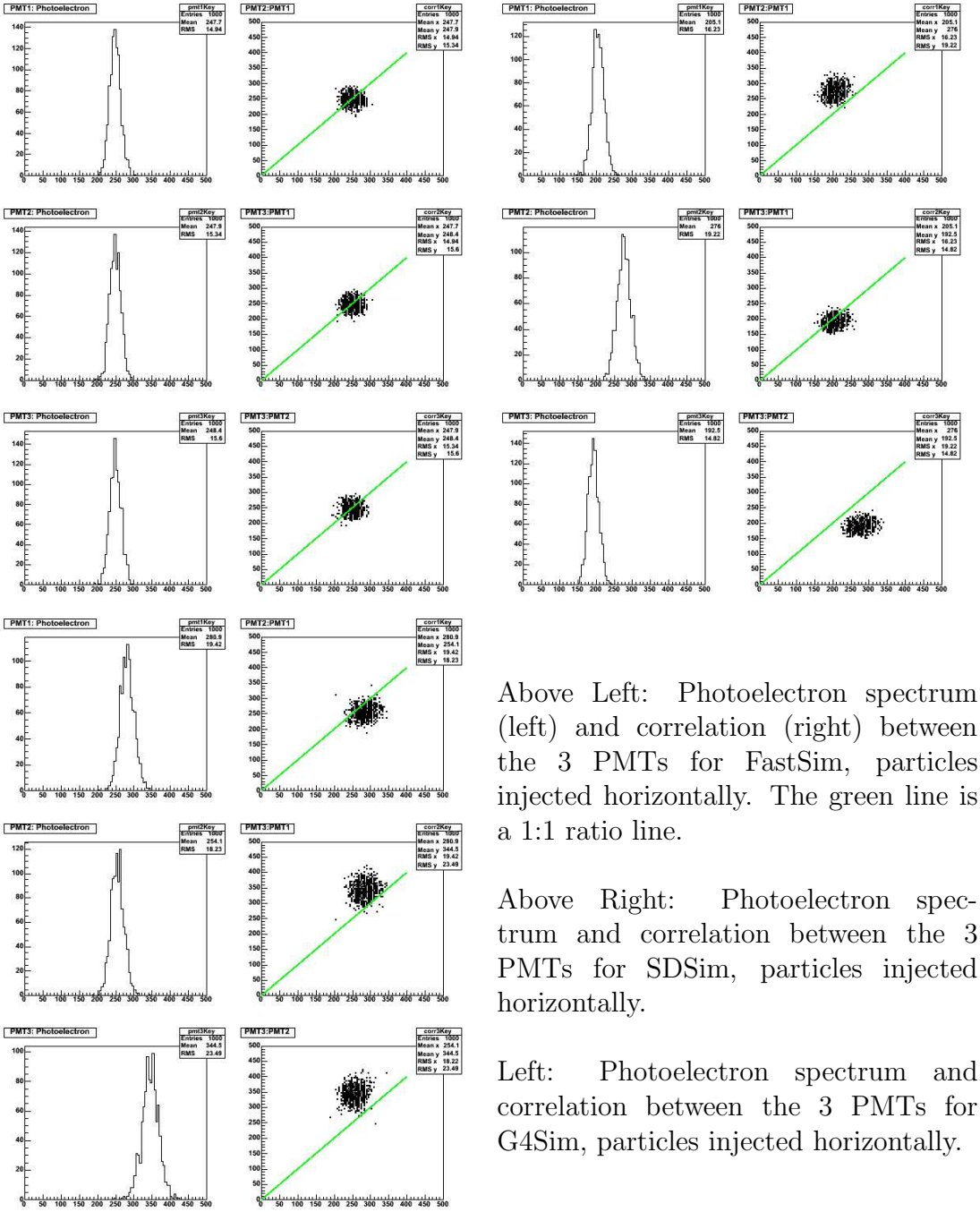


Above Left: Response of FastSim to three particle types injected horizontally at 0.6 m above the ground. The upper plot is the total photoelectrons as a function of incident energy for three particle species.
The lower plot is the fluctuation excess, the fraction of the spread of the fitted mean divided by the square root of the mean.

Above Right: Response of SDSim to three particle types.

Left: Response of Geant4 to three particle types.

Figure 7.8: Response of the three tank simulations to three particle species, mouns, electrons and gammas injected horizontally, as linear plots.



Above Left: Photoelectron spectrum (left) and correlation (right) between the 3 PMTs for FastSim, particles injected horizontally. The green line is a 1:1 ratio line.

Above Right: Photoelectron spectrum and correlation between the 3 PMTs for SDSim, particles injected horizontally.

Left: Photoelectron spectrum and correlation between the 3 PMTs for G4Sim, particles injected horizontally.

Figure 7.9: Correlation between response of the three phototubes in each tank simulation

7.3.5 Comparison of Single Particle Response

These plots present the response of each simulation by particle type. Figure 7.10 shows the response of each simulation as a function of input energy for electrons (upper left), the response to muons (upper right), and the response to gammas (lower left). Here again, the inclusion of the dead material of the tank affects the response of Geant4 for electrons, Figure 7.11 upper left. The discrepancy between responses in FastSim and SDSim when compared to G4Sim for both electron and gamma at low energy is a problem since these form the majority of particles at ground level and is important for correct shower energy determination.

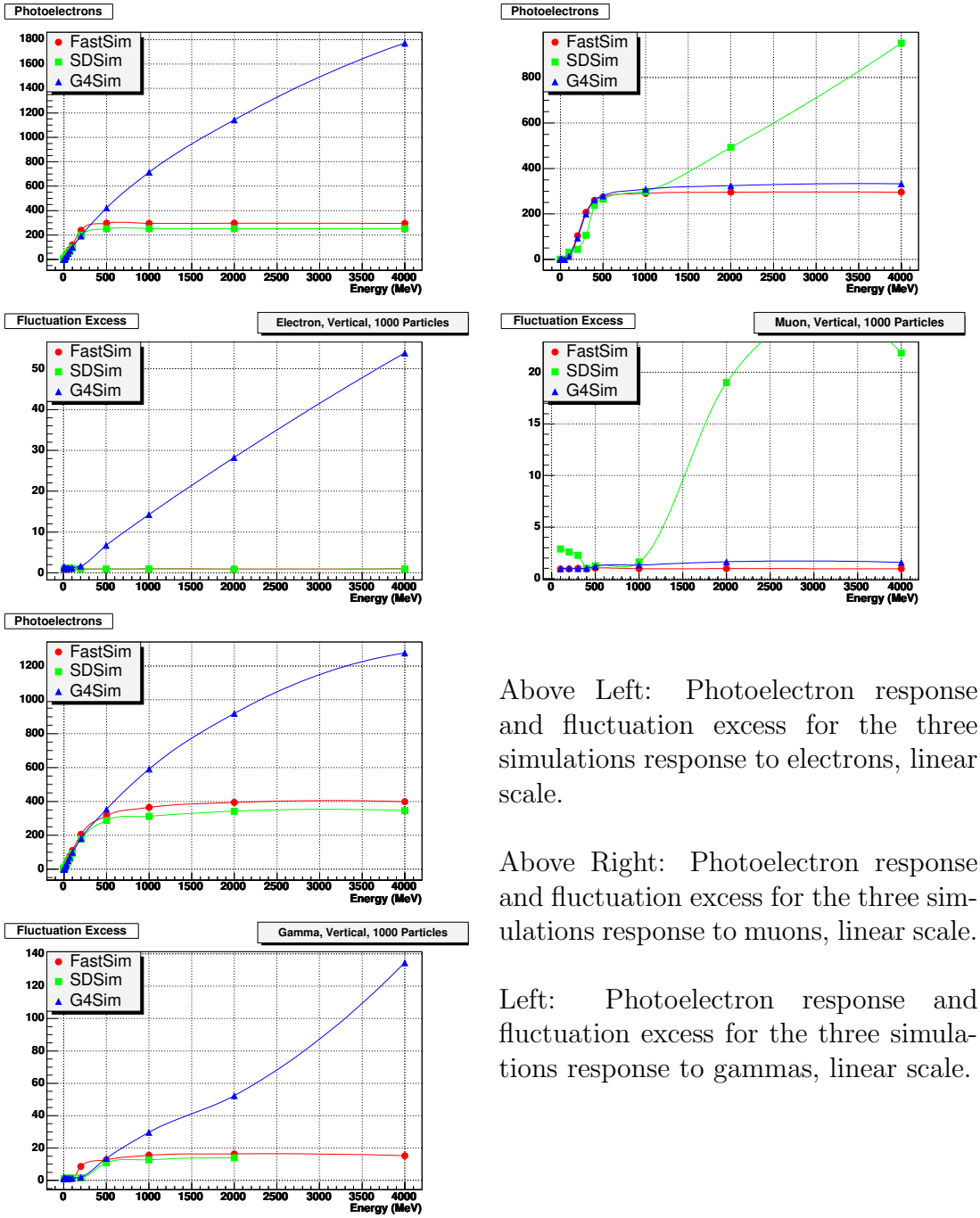
7.3.6 Effect of Muon Decay and Delta Ray in G4Sim

The above Geant4 studies were done with the muon decay process turned off. When this decay channel was enabled, the electron created increased the low energy response as can be seen in Figure 7.12. In addition, the delta ray process was turned on and off with the results shown in Figure 7.13 left. The $\sim$10% difference with delta ray on and off is important for VEM simulation as it directly affects this parameter.

7.3.7 Effect of Gamma Conversion Probability

One major contribution to the difference in tank response to gammas, Figure 7.11, is the finite probability of a gamma to shower within the tank. This can be determined by the number of gammas that do not interact in the tank, giving a response of zero photoelectrons as shown in Figure 7.5. For Fast and SDSim, the probability rises quickly, reaching 100%, whereas G4 reaches $\sim$90%, as shown in Figure 7.13 right. Again, this discrepancy in conversion probability will have

consequences for shower energy determination since the change in response is steep at the energies of shower particles.

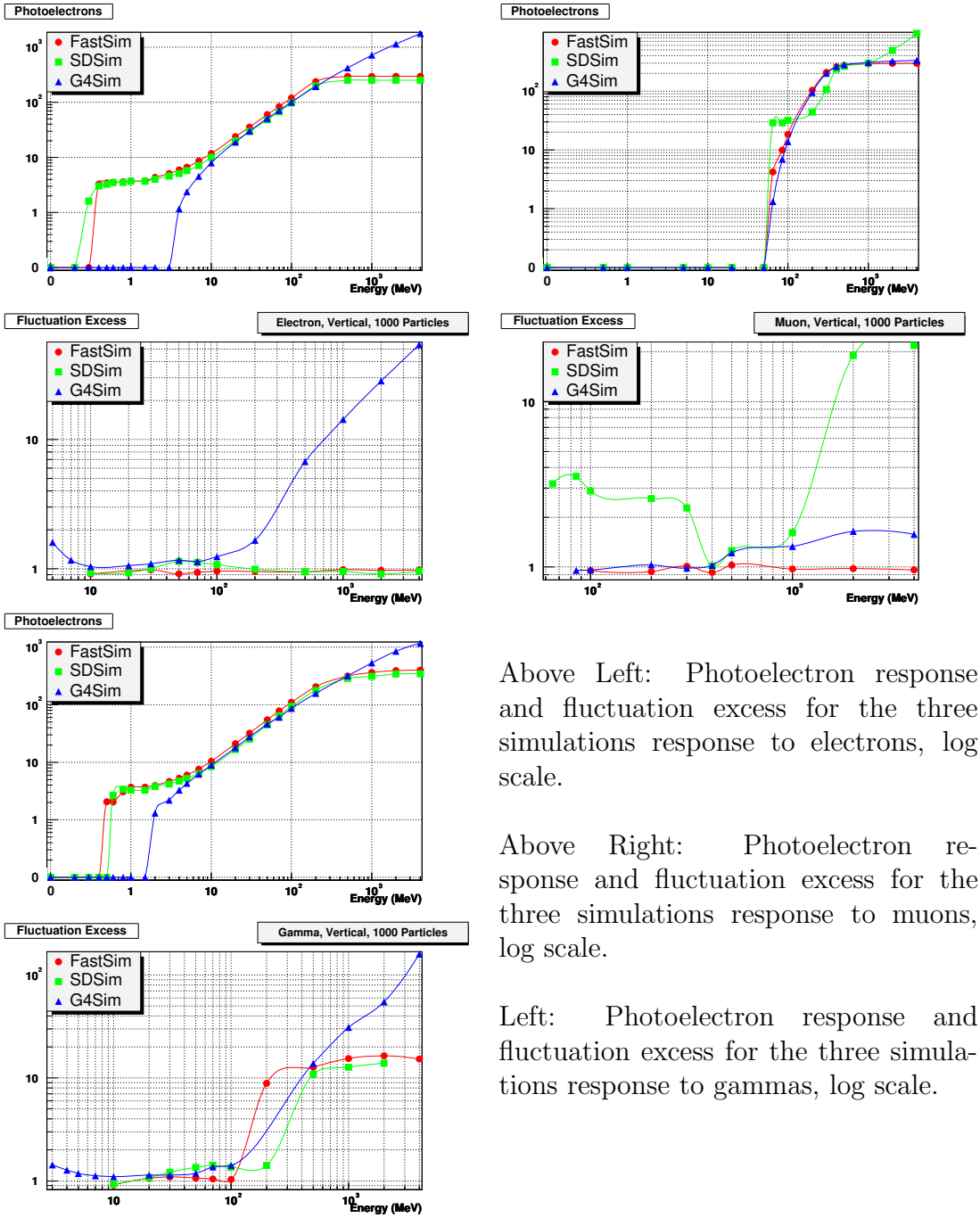


Above Left: Photoelectron response and fluctuation excess for the three simulations response to electrons, linear scale.

Above Right: Photoelectron response and fluctuation excess for the three simulations response to muons, linear scale.

Left: Photoelectron response and fluctuation excess for the three simulations response to gammas, linear scale.

Figure 7.10: Comparison of the three tank simulations by three particle species, electrons, muons and gammas injected vertically, as linear plots.



Above Left: Photoelectron response and fluctuation excess for the three simulations response to electrons, log scale.

Above Right: Photoelectron response and fluctuation excess for the three simulations response to muons, log scale.

Left: Photoelectron response and fluctuation excess for the three simulations response to gammas, log scale.

Figure 7.11: Comparison of the three tank simulations by three particle species, electrons, muons and gammas injected vertically, as log plots.

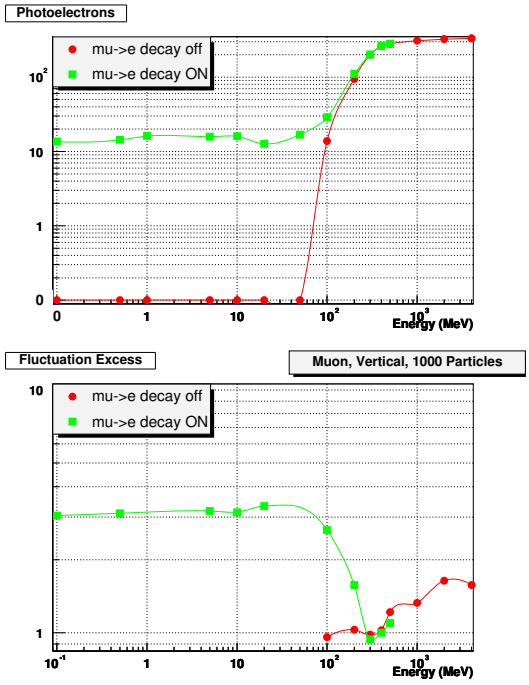


Figure 7.12: Photoelectrons and fluctuation excess in Geant4 with and without muon decay

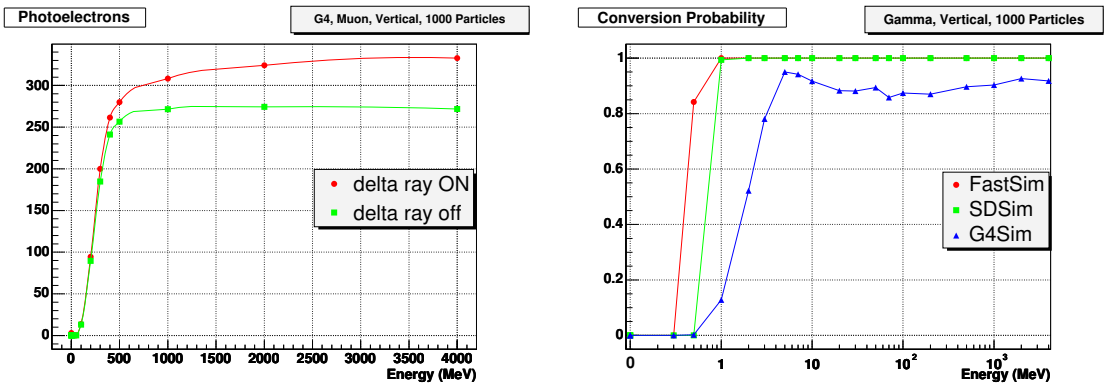


Figure 7.13: Left: Photoelectron response in Geant4 with delta ray on and off. For energies less than those shown, effect was minimal.
Right: Conversion probability of gammas in Geant4. This was determined by the fraction of zero PE events vs the total number of events.

7.3.8 Photoelectron Trace

The Event dump module can also output the photoelectrons as a function of arrival time, these are shown in Figures 7.14 to 7.17. For both G4 and SDSim, there is an initial spike which may be due to the photons from the first reflection and a long tail from the subsequent reflections within the tank. If the trace is rebinned to 25ns (the time division of the real FADC), a pulse shape similar to those observed on real tanks is seen with the characteristic decay time of ∼65ns, see Figure 7.15. In the case of FastSim, the photoelectron trace is generated from a XML table of the cumulative probability of emission, so the pulse shape is arbitrary and entirely configurable, Figure 7.17.

7.4 Conclusion

7.4.1 Summary

Using the DPA framework, we exercised the three available tank simulations and found differences in performance. These differences and bugs in the simulations have been reported to the module authors and are being fixed. The major differences between simulations is due to the sophistication of the physics processes being modeled and inclusion of the plastic tank material.

7.4.2 Future Prospects

The tank simulation is only the first step in understanding the behavior of the ground array. Future work would include completion and testing of PMT simulation and electronics simulation. PMT simulation is required to provide a realistic signal to the electronics including artifacts such as afterpulsing, non-linearity, sat-

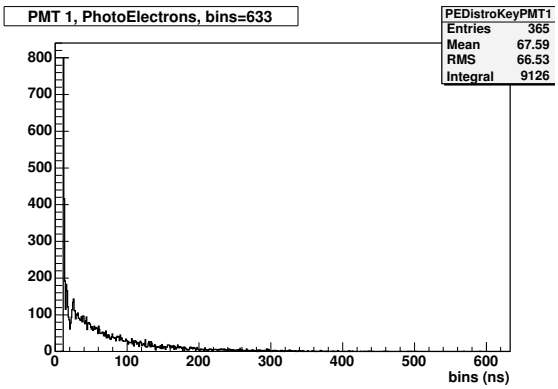


Figure 7.14: Geant4 photoelectron time trace

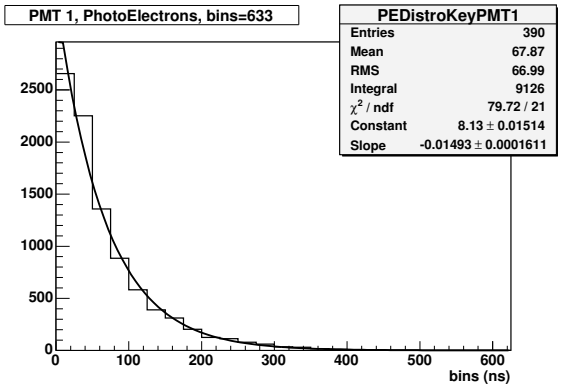


Figure 7.15: Geant4 photoelectron trace rebinned to 25ns, fit to slope gives 66ns

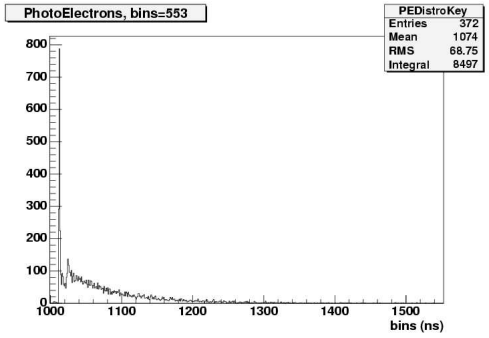


Figure 7.16: SDSim photoelectron time trace

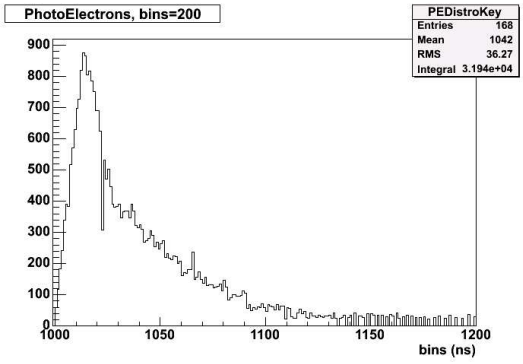


Figure 7.17: FastSim photoelectron time trace

uration, noise, power supply and temperature effects. The electronics simulation would include behavior such as filter effects, amplifier non-linearity, saturation, noise and quantization effects. The great advantage of using the DPA framework is the ability to test each module independently and make corrections and then the entire simulation chain.

	Previously	Now
Response to EM	∼10%	∼5%
Muon Hump/ VEM Ratio	∼10%	∼5%
Dynode / Anode Ratio	∼10%	∼5%
Non-Linearity	∼5%	∼5%
Waterdepth, Area	∼5%	∼5%
Thermal Effect	∼3%	∼3%
Total Systematic Error	∼19%	∼10%

Table 7.6: Detector related uncertainties

7.4.3 Muon Hump/VEM Study

Initial work checking this vital calibration parameter has been done by Gonzalo Rodriguez-Fernandez in an upcoming gapnote [16], but a full comparison of Geant4, lookup table and real data has yet to be done.

7.4.4 Evaluation of Detector Related Systematic Error

Current studies of detector related effects are seriously hampered by the slowness of the Geant4 simulation. The energy determination of a shower is critically dependent on the tank's response to low energy electrons and gammas since these form the majority at ground level. As shown in the results, the conversion probability of gammas and the effect of the plastic of the tank on it's response to electrons needs to be well characterized in simulation. Studies of various fitting algorithms and functions need an accurate tank simulation since the results are heavily affected by the sparseness of the array such that the quality of the signal from each tank, especially near the core, is important.

The tank simulation is important in understanding the effects of each of the uncertainties listed in Table 7.6. These numbers are current best estimates of the uncertainties but future work using a lookup table simulator will analyse the

propagation of a change in each parameter to the final energy determination. This involves a two step process, the first step is an analysis of the Geant4 simulation with these parameters varied, and then transferring the response to a lookup table simulator which will be used with the simulated showers.

CHAPTER 8

Small Scale Clustering Analysis

In 1996, AGASA announced the discovery of clustering of ultra-high energy cosmic rays: 3 doublets were found in their data of 36 events with energies greater than 40 EeV [19]. A cluster was defined as either a pair of events (doublet) or three events (triplet) separated by 2.5 degrees. This value of angular separation was justified as the angular resolution of the AGASA array (1.6°) times $\sqrt{2}$. Later, in 1999 after accumulating 47 events with energies greater than 40EeV, they found a triplet [20]. This was soon changed in 2000 with the addition of two doublets in a total of 58 events [18], see Figure 8.1. One of these doublets was a consequence of including an event with an energy of 38.9 EeV. Most recently, [21] they have reported with 59 events the same clusters of 1 triple and 5 doublets with a combined chance probability of 10^{-4}. However, as discussed by Finley and Westerhoff in [22], this significance was not properly confirmed by either using an independent data set or by calculating the statistical penalty associated with scanning to find the point of highest significance. They found that if this was done, the significance drops to 0.003. Recently, HiRES have done a similar analysis and found no significant clustering [23]. In addition, others have found that both AGASA and HiRES are consistent given a density of source distribution and the statistics of the experiments' event count. At present, there does not appear to be statistically significant clustering in the northern sky. To determine if a clustering signal is statistically significant, it is important to sufficiently cover

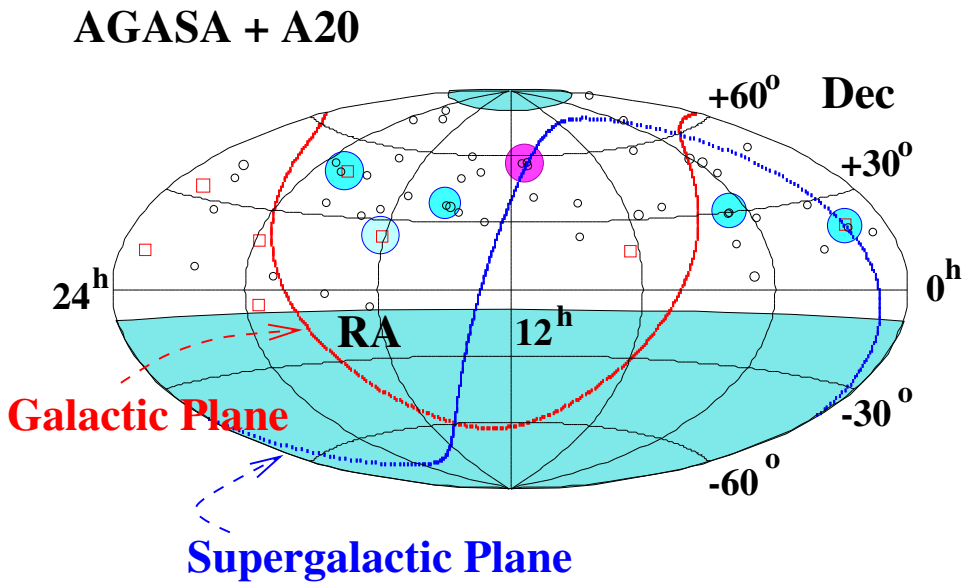


Figure 8.1: AGASA sky map in equatorial coordinates from [18]. Small circles and squares represent events with energies 40–100 EeV and $>$ 100 EeV respectively. Shaded areas are not visible to AGASA

the 'phase space' of the data-what minimum energy threshold to use and what angular separation is considered a correlation. In previous studies of clustering, the angular separation between events, minimum event energy, and both were varied to maximize the signal, but this was not done systematically. Since a priori it is difficult to predict the amount of deflection suffered by a charged cosmic ray on its way to earth, and this bending is dependent on the energy of the particle, it is necessary to scan over a wide range of energy thresholds and angular separations. It is then required to take a penalty in significance for performing this scan. For a neutral particle, the maximum significance should occur at an angular separation that is the resolution of the detector, thus a determination of resolution is needed.

8.1 Methods of Analysis

A search for clustering among the highest energy cosmic rays will yield a significance which is dependent on the minimum energy E_c which defines the data set and the maximum angular separation θ_c which defines a pair. Ideally E_c and θ_c should be determined a priori from a theoretical model, to be tested, of cosmic ray source distributions and Galactic and extragalactic magnetic fields. But at present, uncertainty in extragalactic magnetic fields make it difficult to create accurate models. Previously, AGASA and others have scanned over a range of energies, angular separations and both simultaneously to identify E_c and θ_c that maximize the clustering signal. However, the true significance of this signal must be determined by performing identical scans over simulated data sets for many trials ($\sim 10^6$). For our studies, we used a method developed by Tinyakov and Tkachev, described in detail in [24], and used by Finley and Westerhoff to analyze the AGASA data. In short, for analysis of N cosmic ray data, we would like to first determine E_c and θ_c through direct scanning over minimum energies and angular separations. However, in practice instead of scanning directly over energy thresholds, we rank the events by energy and scan over events n. That is, for each value of n and θ, we count the number of pairs n_p separated by less than θ. To determine E_c and θ_c , we first generate a large number ($n_{mc} \sim 10^5$, typically one order of magnitude greater than the minimum chance probability expected, say $1{:}10^4$) simulated data sets with the same exposure as the detectors, and record at each n and θ the number of instance (n_i) that the simulated sets observe n_p or more pairs. The probability of observing n_p or more pairs at each $N_c$and θ_c will simply be:

$$P_{data}(n, \theta) = \frac{n_i}{n_{mc}}.$$

For some combination of $N_c$and θ_c , P_{data}has a minimum: $P_{min} = P_{data}(N_c, \theta_c$). This identifies the location in the scan of the strongest potential clustering signal. To determine the true significance of this signal, one must perform the scan over N_{mc} Monte Carlo simulated data sets ($N_{mc} \sim 10^5$), then identify the minimum probability $P^i_{min} = P^i(N^i_C, \theta^i_C)$ for each trial and counting the number of trials N^*_{mc} for which $P^i_{min} < P_{min}$. Finally, the chance probability P_{chance} of observing P_{min}in the scan is evaluated as:

$$P_{chance} = \frac{N^*_{mc}}{N_{mc}}.$$

8.2 Verification of Analysis

8.2.1 AGASA Auto-correlation

To test the validity of our programming, we reproduced Finley and Westerhoff's analysis on the 57 AGASA events with energy ¿ 40 EeV listed in [18] (we did not include the event below 40 EeV). When we generated the Monte Carlo simulated events for determining the probabilities, we assumed a uniform exposure of AGASA in right ascension, and a distribution in declination ω described by:

$$\omega(\delta) \sim cos(a_0)cos(\delta)sin(\alpha_m) + \alpha_m sin(a_0)sin(\delta) \tag{8.1}$$

where α_m is given by

$$\alpha_m = \begin{cases} 0 & if\ \xi > 1 \\ \pi & if\ \xi < -1 \\ cos^{-1}\xi & otherwise \end{cases}$$

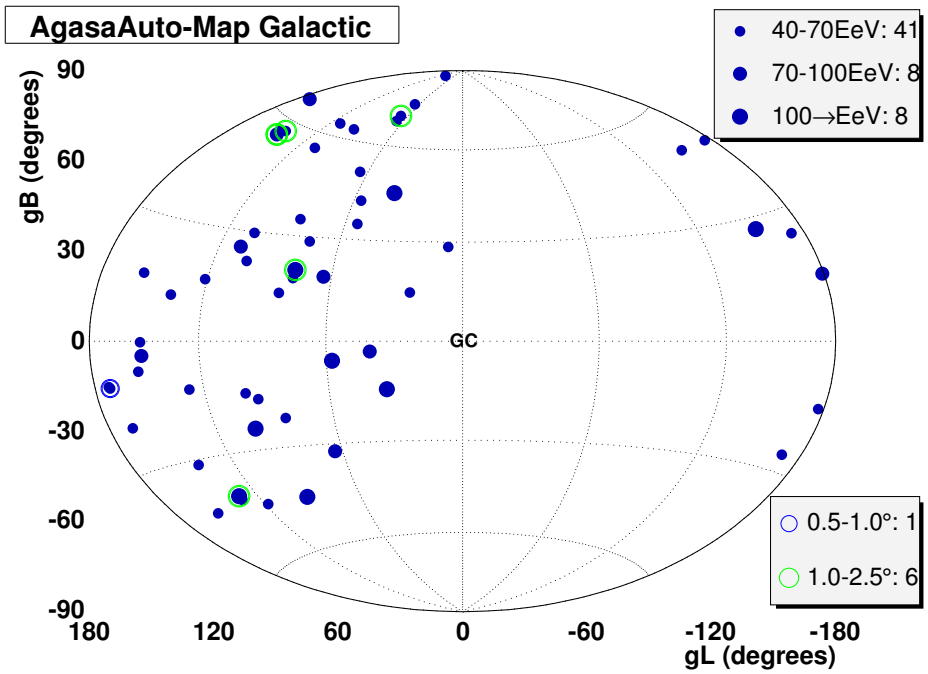


Figure 8.2: Our plot of AGASA data in galactic coordinates. Galactic latitude increases to the left, galactic center is at the center of the map. Size of marker indicates energy of event, size and color of open markers indicate angular separation between correlations. Note that correlation markers are not to scale.

and

$$\xi \equiv \frac{cos(\theta_m) - sin(a_0)sin(\delta)}{cos(a_0)cos(\delta)}$$

with a_0 , set to the geographical latitude of AGASA, to be 36°, and Θ_M, the maximal zenith angle to be 45°. When this function is computed, the theoretical exposure of AGASA is obtained, Figure 8.3. We then scanned over angular separations from 0° to 10° in increments of 0.1° and over the 57 events in order of rising energy. The plots in Figure 8.4 from [22] show $P_{min} = P_{data}(N_c,\theta_c) = 8.4 \times 10^{-5}$, where N_c= 36 highest-energy events (energy threshold is 4.89 $\times 10^{19}$eV), and $\theta_c = 2.5°$.

Similarly, the results obtained in our reproduction, Figure 8.5, show P_{min}= 8 $\times 10^{-5}$, where N_c= 35 highest-energy events (energy threshold is $\sim 4.9 \times 10^{19}$eV),

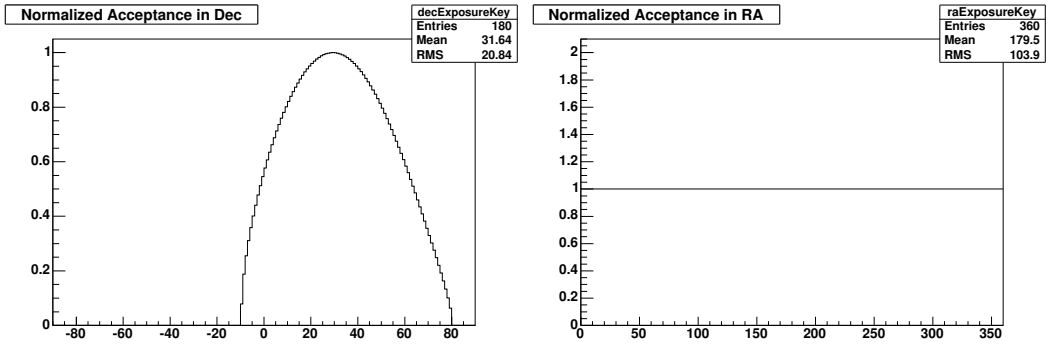


Figure 8.3: On the left, the computed AGASA exposure in declination, on the right, in right ascension

and $\theta_c = 2.4°$. As mentioned earlier, to evaluate the significance of P_{min}, one has to perform the identical scan over simulated AGASA data sets and count how many simulated sets have $P_{min}^{MC} < P_{min}$ of (8.4×10^{-5}). Finley and Westerhoff found that 3475 out of 106 simulated sets meets this condition, implying a chance probability P_{chance} of 0.3.

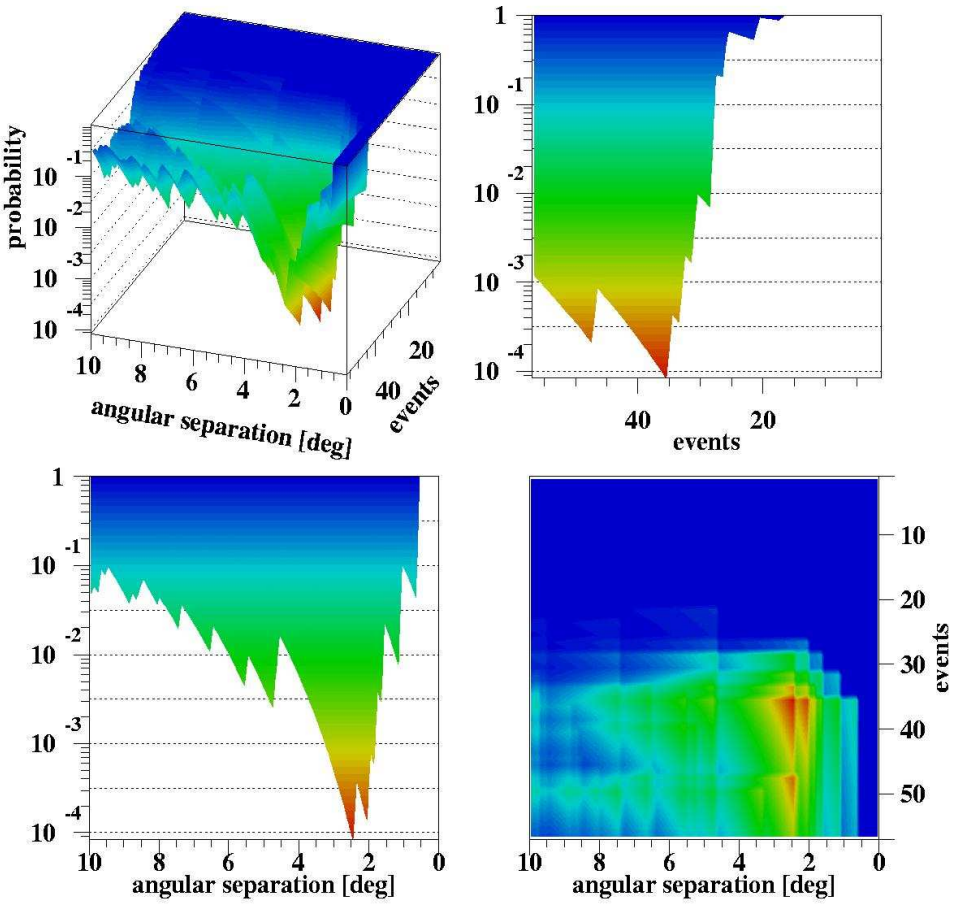


Figure 8.4: Results of Finley and Westerhoff's analysis of AGASA auto-correlation. Maximum significance is at $N_c = 36$ and $\theta_c = 2.5°$.

8.2.2 HiRES Auto-correlation

The same autocorrelation analysis was performed on HiRES data (Fig. 8.6), which consists of 271 stereoscopic events with reconstructed energy $E > 10^{19}$eV [25]. However, the energies of the events are not published, so one can only perform a scan over angular separation. Figure 8.7 shows the exposure of HiRES experiment, which shows a non-isotropic distribution in right ascension, since HiRES, being a fluorescence experiment, operates only on clear moonless nights. This, in combination with the earth's orbit around the sun, gives a non-uniform

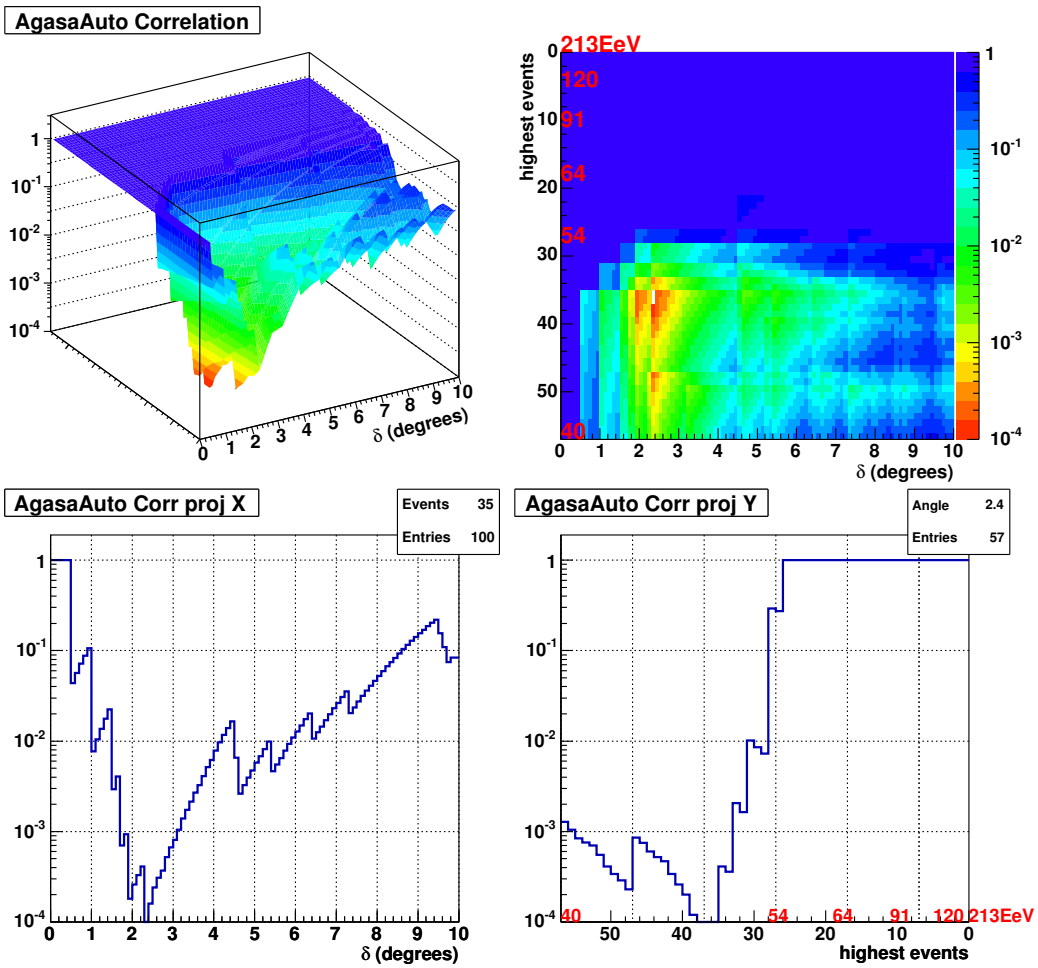


Figure 8.5: Results of our analysis of AGASA auto-correlation. Apart from the reversal of the plots, this result essentially reproduced that shown in Figure 8.4. Greatest significance is at $N_c = 35$ and $\theta_c = 2.4°$.

exposure in right ascension. In our analysis, we approximated the HiRES exposure from the data points in this figure. We then generated our MC data sets according to this exposure, and again scanned over angular separations from 0° to 10° in increments of 0.1°. The result of this scan is Figure 8.9 and it shows that $P_{min} \sim 0.3$, where $\theta_c = 1.2°$. In Figure 8.8 from [26], the HiRES data show a maximum significance at $E_c = 16.9$ EeV and $\theta_c = 2.2°$ of 0.019. Though we cannot directly compare our result with theirs, our one dimensional scan at 10

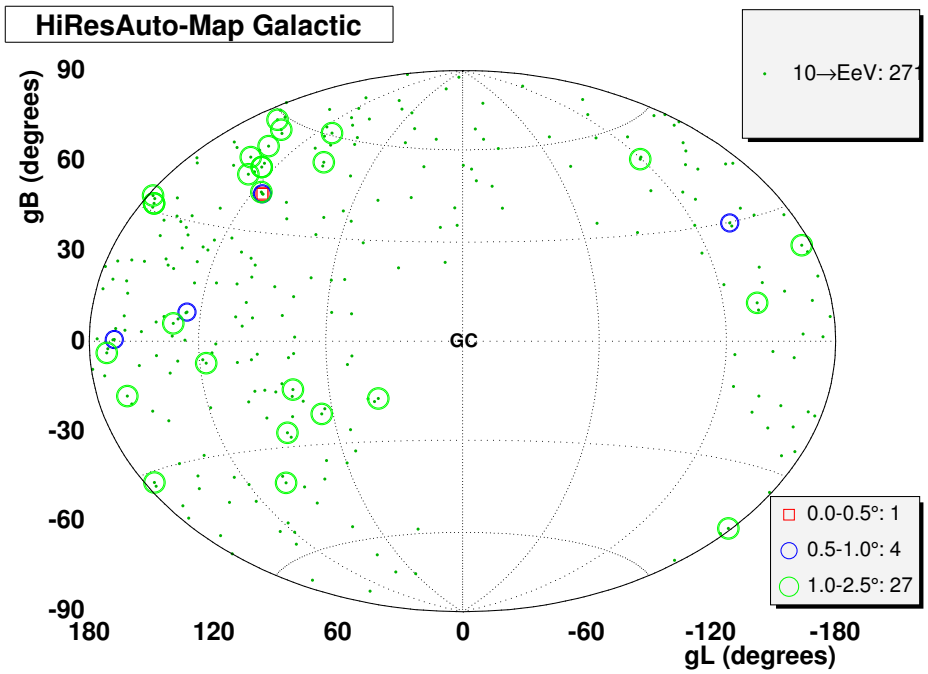


Figure 8.6: Our plot of HiRES data with energy > 10 EeV in galactic coordinates, sixe and color of hollow markers indicate angular separation between correlations.

EeV should be equivalent to a 'slice' of their significance plot at 10 EeV with our angles of highest significance, 1.3° and 2.2° matching those in Figure 8.8.

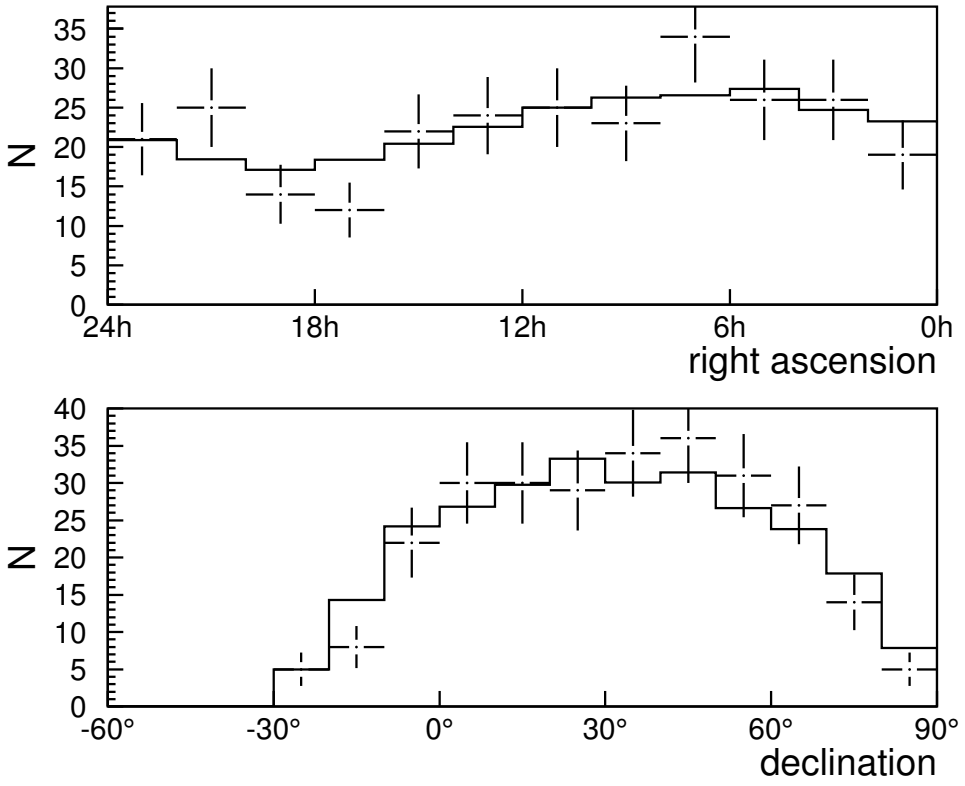


Figure 8.7: Published HiRES exposure from [26], the line is MC computed exposure, the points actual HiRES exposure - these data were used to generate the exposure in this analysis.

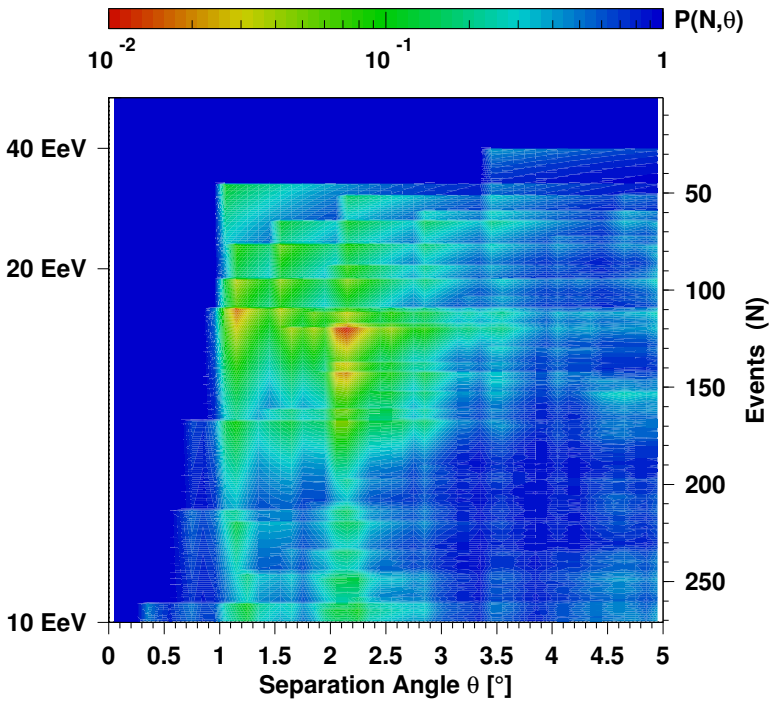


Figure 8.8: 2-D probability scan from [26] showing a maximum significance at E_c = 16.9 EeV and $\theta_c = 2.2°$.

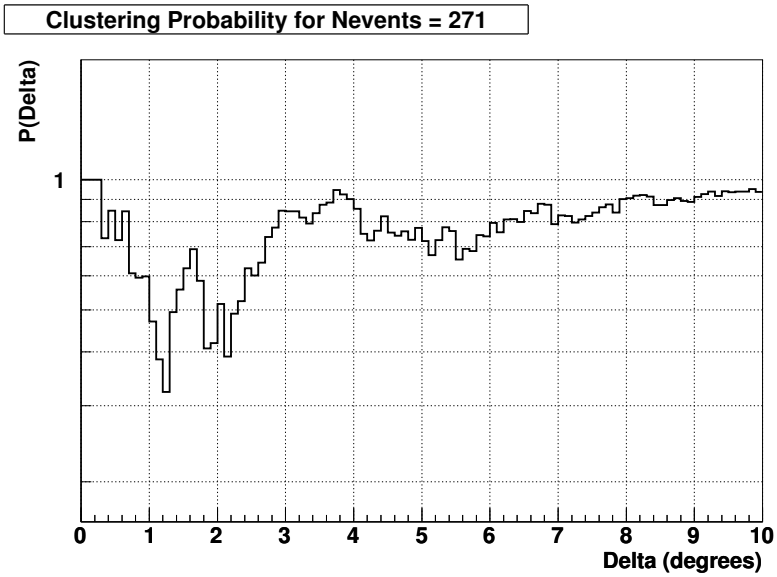


Figure 8.9: Results of HiRES auto-correlation analysis, no high significance is seen. Since we did not have energies for each event, this result cannot be directly compared to the HiRES result, however the points of highest significance match those in Figure 8.8

CHAPTER 9

Auger Clustering Analysis Preliminary Results

9.1 The Auger Dataset

Using the techniques described above, we analyzed the Auger data which were selected as follows:

- Data from the period 1 January to 31 December 2004. The array grew from 190 to 603 tanks during this period (Fig. 9.1).
- Data reconstructed within the DPA framework (v1r0p0) with Joong Lee's SDReconstructorUCLA, for details see [12 FIXME].
- Reconstructed energy (assuming proton and QGSJet) < 10 EeV and with zenith angle cuts of 45, 60 and 75 degrees. For details see [13 FIXME].
- Quality - We required a positive Beta, positive Shower Front Curvature and Positive FADC Risetimes, and one good tank (a tank with unsaturated signal) within 1000 m from the core

With these requirements we obtain the following event distribution:

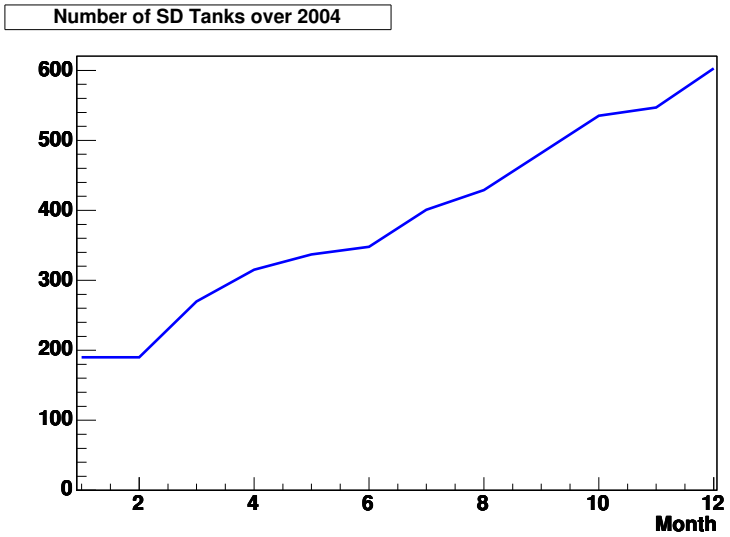


Figure 9.1: Auger tank count as a function of time for 2004

Experiment	Zenith	> 10 EeV	> 20 EeV	> 40 EeV
Auger Jan-Dec 2004	45	254	59	13
	60	453	102	23
	75	639	147	38
HiRES Dec 1999 to Jan 2004	70	271		27
AGASA 1990 to May 2000	45			57

Table 9.1: Event counts for Auger SD data 1 Jan to 31 Dec 2004 at zenith angles upto 45, 60 and 75° and with reconstructed energy > 10, 20 and 40 EeV. For comparison, HiRES [25] and AGASA [18] event counts are shown

9.2 Angular Resolution

Since accuracy of pointing direction is important in clustering studies, a study was performed to determined the Auger SD angular resolution. Simulated showers created by AIRES-QGSJET with a proton primary were reconstructed using the DPA framework with the SDReconstructorUCLA module and the reconstructed direction compared with the input to AIRES, for more details, please see [12 FIXME]. The results can be summarized by the plots in Figures 9.2 and 9.3.

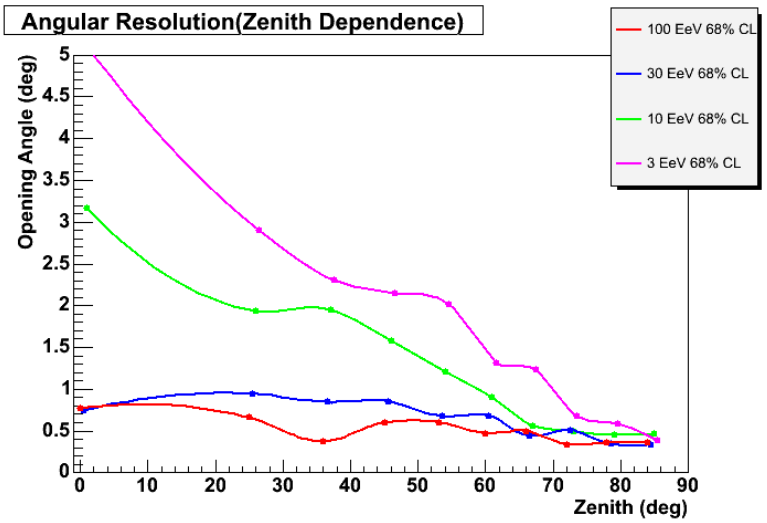


Figure 9.2: Auger SD angular resolution as a function of zenith angle. Note that accurate reconstruction can be done even at the highest zenith angles, for energies >40 EeV the resolution is less than a degree.

In Figure 9.2, it is clear that there is a general improvement of resolution with increasing zenith angle and since the majority of cosmic ray arrival phase space is at larger zenith angles, the resolution is better than 1 degree and approaching 0.5 degree at higher energies for much of our data. In Figure 9.3, there is also an improvement of resolution with increasing energy. Again for events with energy greater than 10 EeV and at larger zenith angles the resolution is a degree or better. An unexpected result of this study was our ability to reconstruct high zenith angle ($> 70°$) events with good angular resolution. At large zenith angles, the shower at the ground is mostly composed of muons which, inducing a fast risetime in the tank, gives a clear arrival time for the particles thus enabling an accurate geometric reconstruction. The energy determination at high angles is still reasonable (Figure 9.4), so data up to 75 degrees are included in this paper.

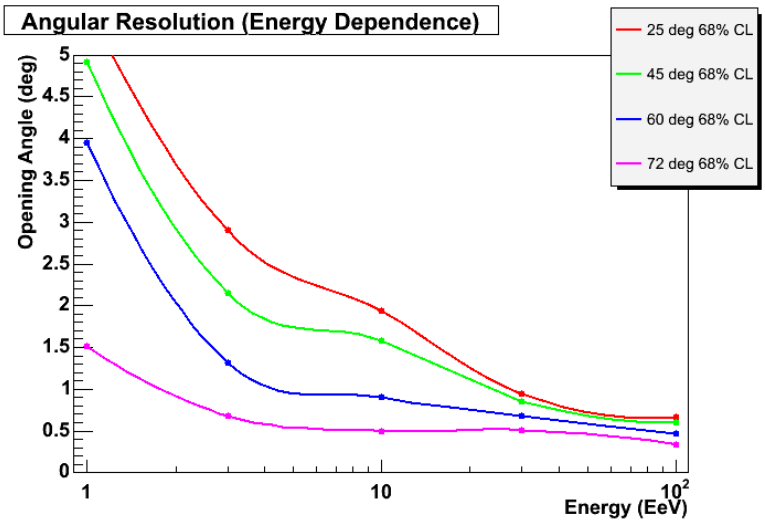


Figure 9.3: SD angular resolution as a function of energy. The opening angle is the difference between the reconstucted direction and the monte carlo input. Note that with energies in excess of 40 EeV, the resolution is better than 1°. These plots courtesy of Joong Lee.

9.3 Sky Maps

To present the Auger data in a consistent manner, we use Galactic Coordinates going positive to the left, on a Hammer-Aitoff projection with the galactic center in the center of the plot. This is an equal area projection which was chosen for ease of comparison with other astrophysical research. Figure 9.5 shows computed coverage for the Auger southern observatory for increasing zenith angles. When compared to the coverage of AGASA and HiRES in Figures 3 and 7 FIXME, it is clear that we would like to use the maximum possible zenith angle to maximize overlap with these experiments. In Figures 9.6 to 9.8, Auger data with a reconstructed energy of greater than 10 EeV are presented. In the top plot, the size and color of the marker indicates the reconstructed energy. In this presentation, the exposure of the detector can be seen as the underexposure in the 45 degree plot at approximately -30 latitude and -60 longitude which becomes over

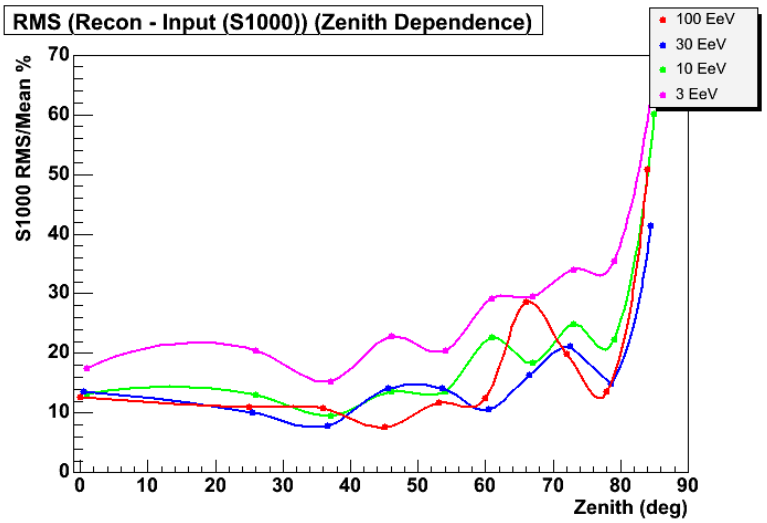


Figure 9.4: Statistical uncertainty in S(1000) as a function of zenith angle. This is an indicator of the uncertainty in the energy determination.

exposed in the 75 degree plot. This feature should be kept in mind when looking at density variations in these plots, and is automatically compensated for in the analyses. In the bottom map of Figures 9.6 to 9.8, the size of the solid marker indicates the event energy and the size and color of the open markers indicates the angular separation of the correlation between the data. The same programs described in Chapter 8 were used to analyze the Auger data for auto-correlations. Note that the size of the hollow marker is not to scale and they are doubled to make them easier to see.

9.4 Auger Auto-Correlation

Though to the eye there appear to be many interesting clusters, the statistical significance as determined by monte carlo indicate otherwise. Correction for exposure was done using the analytical formula described earlier with appropriate values for Auger's latitude and zenith angle acceptance. Though there are more

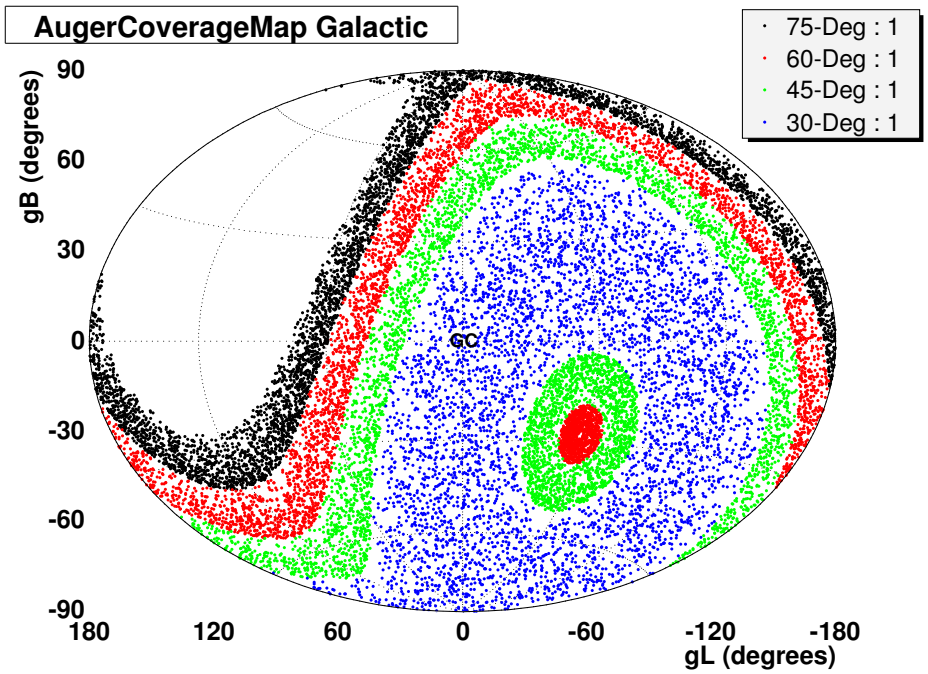


Figure 9.5: Auger coverage map in galactic coordinates for zenith angles less than: 30°(blue), 45°(green), 60°(red) and 75°(black). Each zenith cut includes smaller cuts, e.g. zenith <75° includes all four colored bands

sophisticated methods of determining the exposure (Li-Ma or scrambling), this technique is probably adequate since this analysis is looking for correlations at small angular separation, not trends over large areas of the sky. In Figures 9.9, 9.10 and 9.11, the upper left plot is a 3-D presentation, the x-axis is the angular separation of correlated points (θ_c), the y-axis is the energy threshold (N_c and E_c) and the z-axis the probability, the deeper the point the higher the significance. The upper right plot is a 'top view' of the 3-D with the color showing increasing significance from blue to red. Bottom left and right plots are projections onto the x and y-axes of the 3-D plot through the point of maximum significance, the left is projected onto the angular separation axis, the right is projected onto the event cut axis. This maximum is the raw probability and needs to be confirmed by either an independent data set or by calculation of the

penalty factor, described in Section 8.1 and demonstrated in Section 9.5. The above analysis was repeated for Auger data with energies in excess of 40EeV. This energy threshold is more appropriate for small angle correlations as magnetic deflection will be less. Figures 9.12 and 9.13 show the results of the same analysis as Figures 9.9 to 9.11 except that the minimum energy was set at 40 EeV. The analysis performed with a higher minimum energy has few points and as can be seen in Figure 9.12, no significant auto-correlation. This makes sense since these scans are a 'magnification' of the highest energy regions of Figures 9.9 to 9.11.

9.5 Computation of True Significance

As mentioned earlier, the two dimensional scan technique for auto-correlation requires that either an independent data set be used to confirm the significance of any finding, or that a penalty factor be computed to determine the true significance. If we had observed high significance the latter test would have been computationally intensive but since we had not, the exercise was straightforward. To compute the corrected probability, 1000 simulated data sets were created each with a random distribution of points on the sky following the computed exposure of Auger using Eqn 8.1. The number of points in each simulated data set was the same as the real data and this was done for all combinations of zenith angles (45, 60, 75) and energy thresholds (10 EeV, 40 EeV). Only 1000 sets were needed as the highest significance seen was $> 10^{-3}$. To save additional time, the number of simulated skies in the analysis program was reduced from 10^5 to 10^4. This low number of simulated data sets was adequate for this test as the dependence of the probability on the number of sets reaches an asymptotic limit as can be seen in Figure 9.14.

Each of these simulated data sets was run through the same 2-D scanning

	Energy	> 10 EeV	Energy	> 40 EeV
Zenith	Raw	Corrected	Raw	Corrected
$< 45°$	0.046	0.85	0.49	0.64
$< 60°$	0.034	0.83	0.32	0.76
$< 75°$	0.003	0.31	0.097	0.62

Table 9.2: Raw and corrected probabilities of Auger auto-correlation analyses. For the lowest probability, reduction of significance is two orders of magnitude.

procedure as the real data and the minimum probability was recorded. The fraction of the number of times that the probability from a simulated set was lower than that from real data, normalized by the number of simulated sets, gives the corrected probability. For instance, in the case of 10 EeV energy threshold and 75° zenith angle, the simulated data set gave a chance probability smaller than 0.003 313 times out of 1000 trials, yielding a corrected chance probability of 0.31. Table 9.2 lists the results of this analysis. It is clear that once the true probability is computed, even the interesting low probability point in the 75 degree data is not significant.

9.6 Correlation with BL Lacertae Objects

9.7 Correlation with Previous Experiments

9.8 Discussion

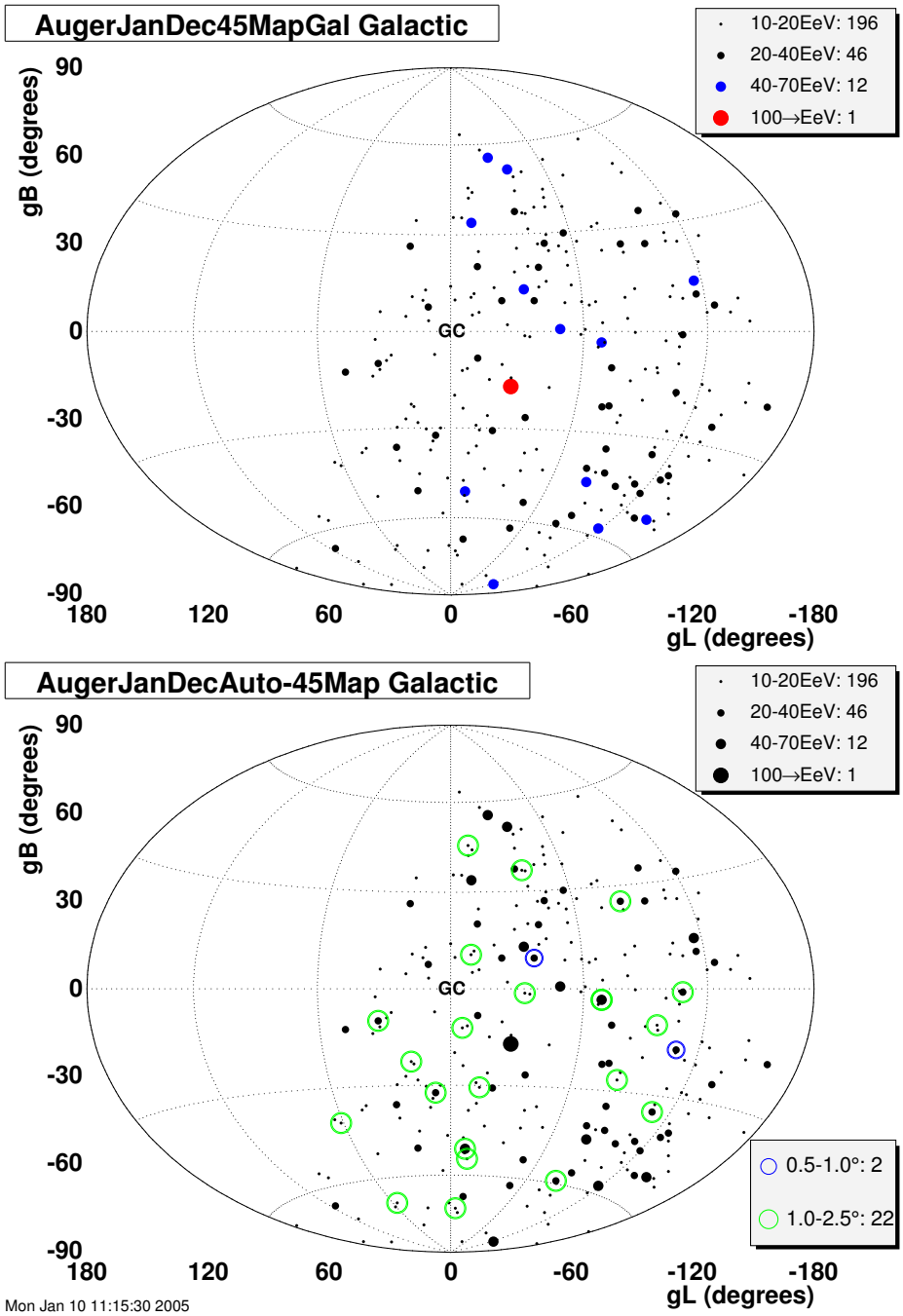


Figure 9.6: Auger data with energy >10 EeV and zenith < 45° plotted in galactic coordinates. Colors and size of marker indicate reconstructed energy. Note underexposure at -60°, -30° in the 45° plot becomes overexposed at the same location in the 75° plot, Figure 9.8

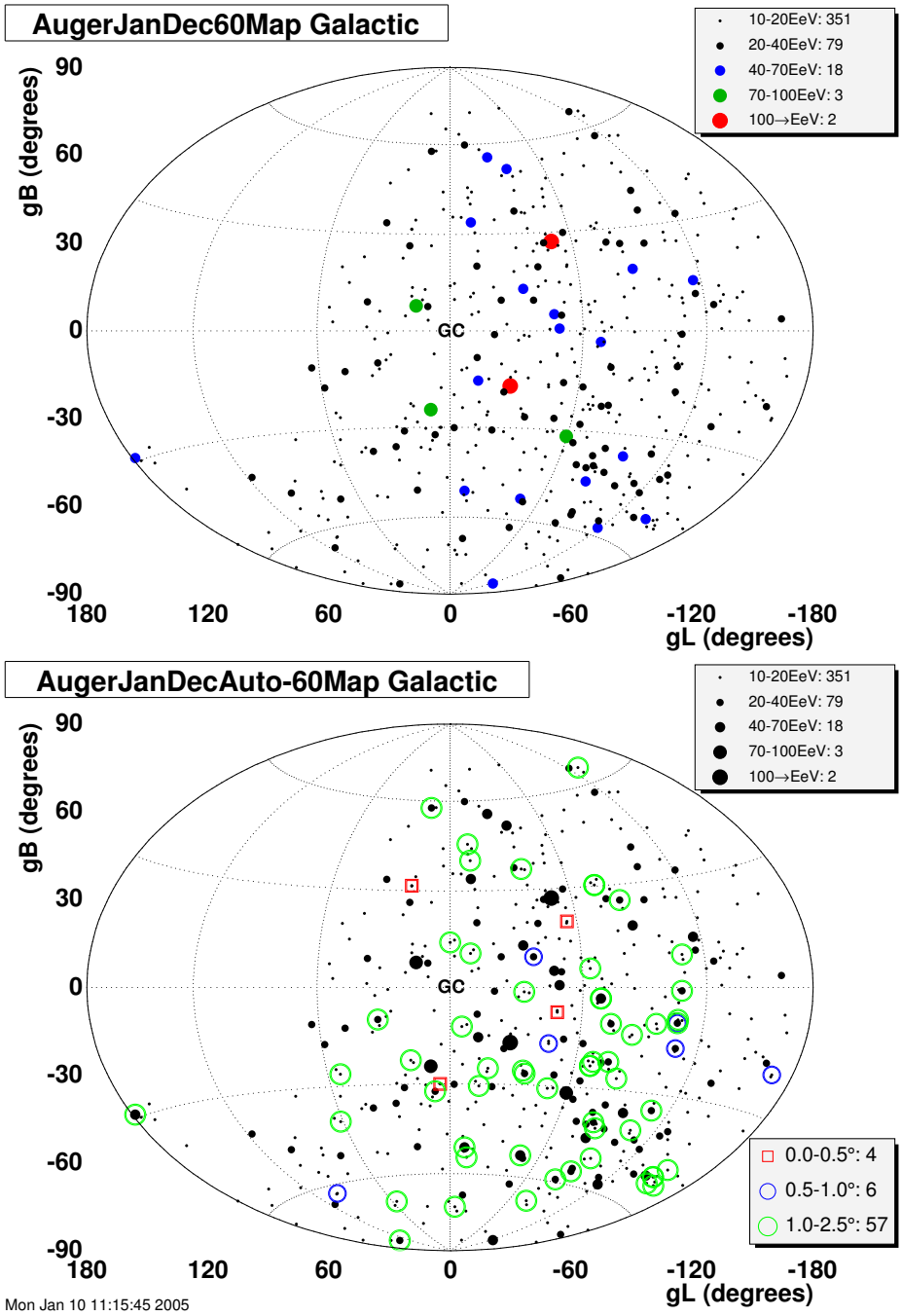


Figure 9.7: Auger data with energy >10 EeV and zenith $< 60°$ plotted in galactic coordinates. Colors and size of marker indicate reconstructed energy.

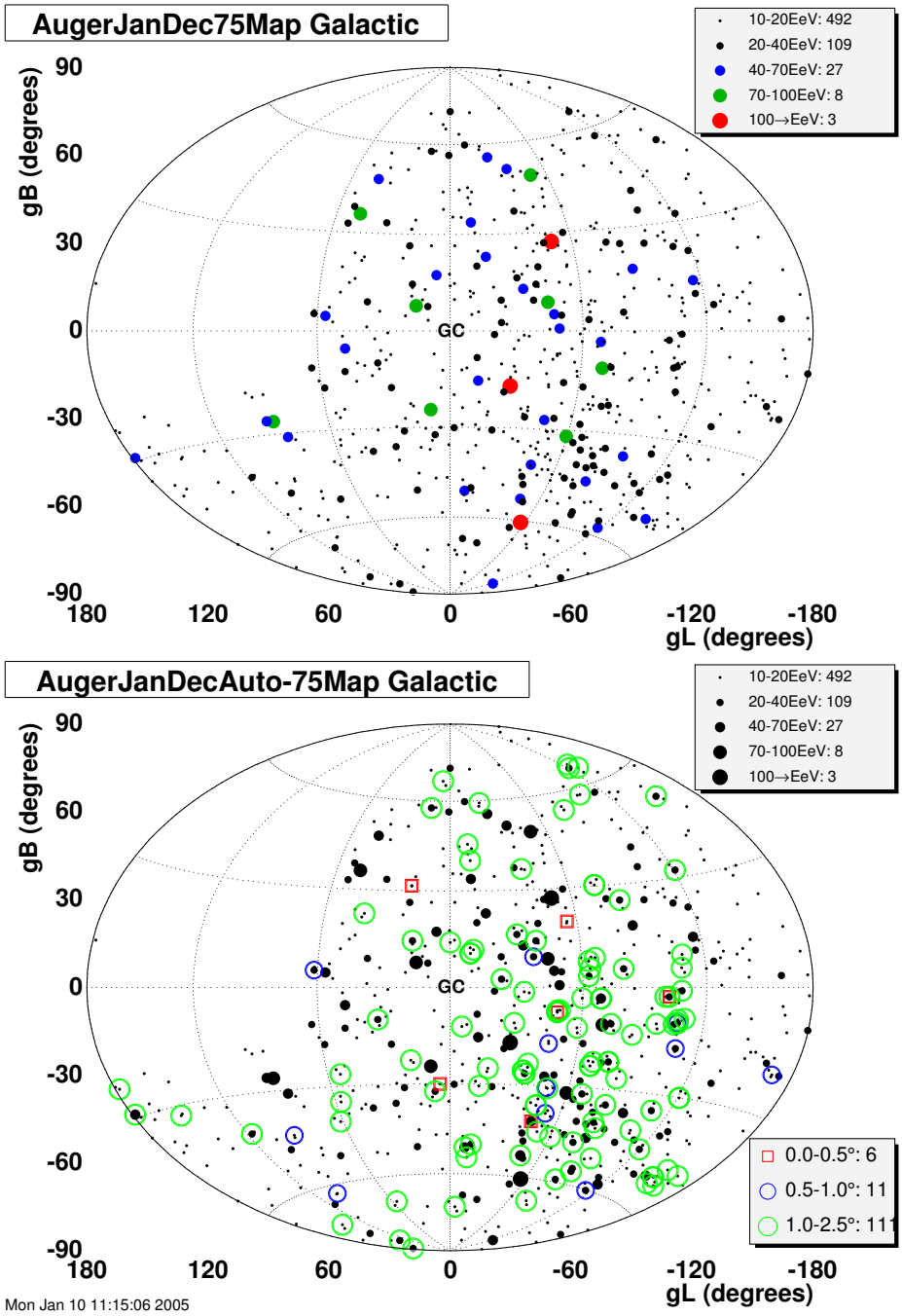


Figure 9.8: Auger data with energy >10 EeV and zenith < 75° plotted in galactic coordinates. Colors and size of marker indicate reconstructed energy.

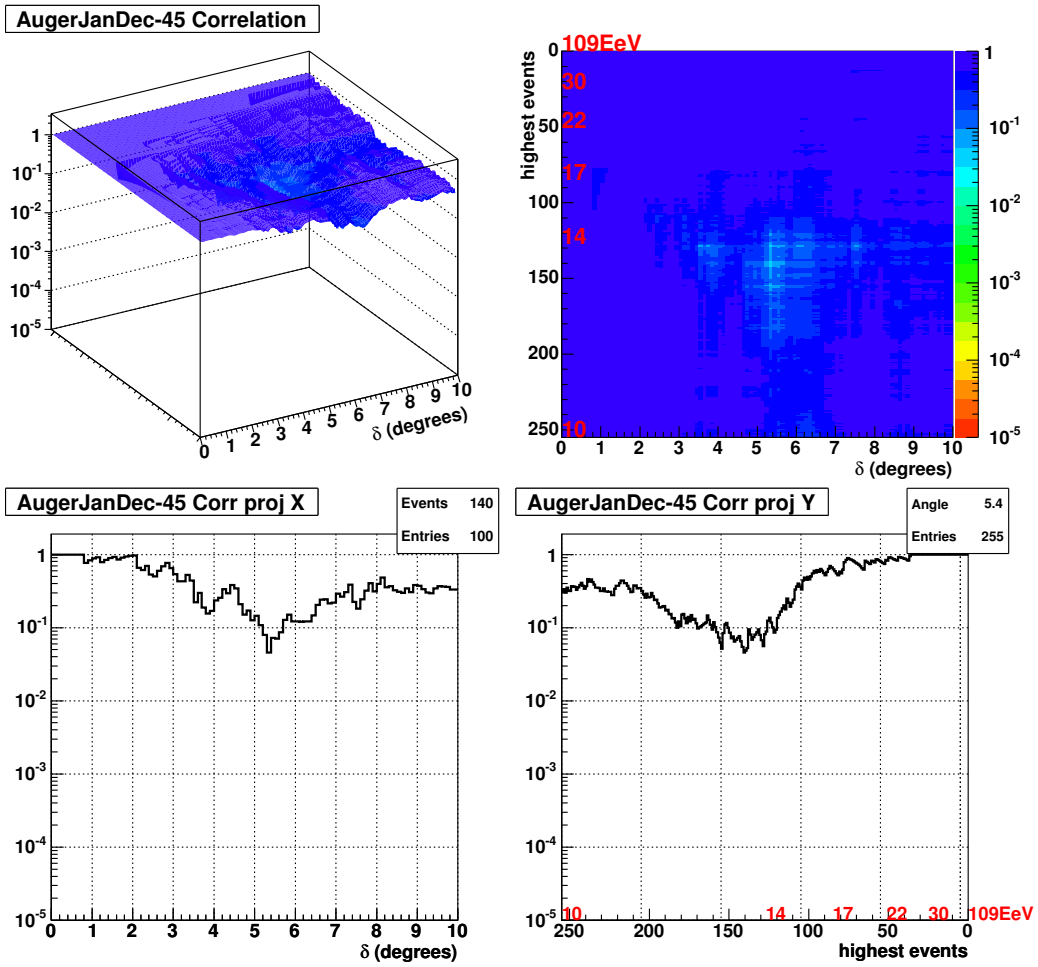


Figure 9.9: Two dimensional scan for significance of auto-correlation of Auger data up to 45°. Maximum significance is at $N_c = 140$ and $\theta_c = 5.4°$.

Top Left: 3-D presentation, x-axis is separation of correlated points in degrees, y-axis is energy threshold, z-axis is probability.

Top Right: top view of 3-D plot, on y-axis, red numbers indicate energy threshold in EeV, color indicates probability.

Bottom Left: Slice of 3-D plot through point of maximum significance projected onto angular separation axis.

Bottom Right: Section of 3-D plot through point of maximum significance projected onto energy threshold axis.

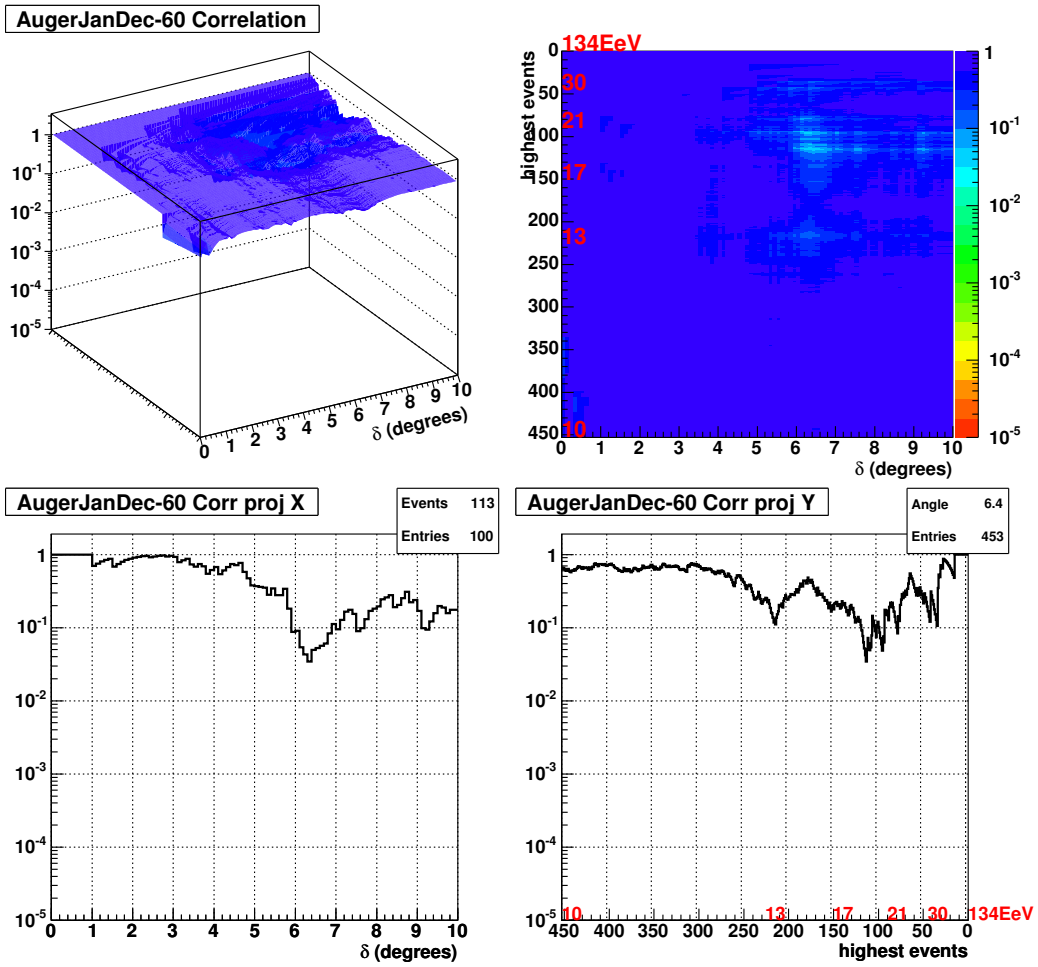


Figure 9.10: Two dimensional scan for significance of auto-correlation of Auger data up to 60°. Maximum significance is at $N_c = 113$ and $\theta_c = 6.4°$.

Top Left: 3-D presentation, x-axis is separation of correlated points in degrees, y-axis is energy threshold, z-axis is probability.

Top Right: top view of 3-D plot, on y-axis, red numbers indicate energy threshold in EeV, color indicates probability.

Bottom Left: Slice of 3-D plot through point of maximum significance projected onto angular separation axis.

Bottom Right: Section of 3-D plot through point of maximum significance projected onto energy threshold axis.

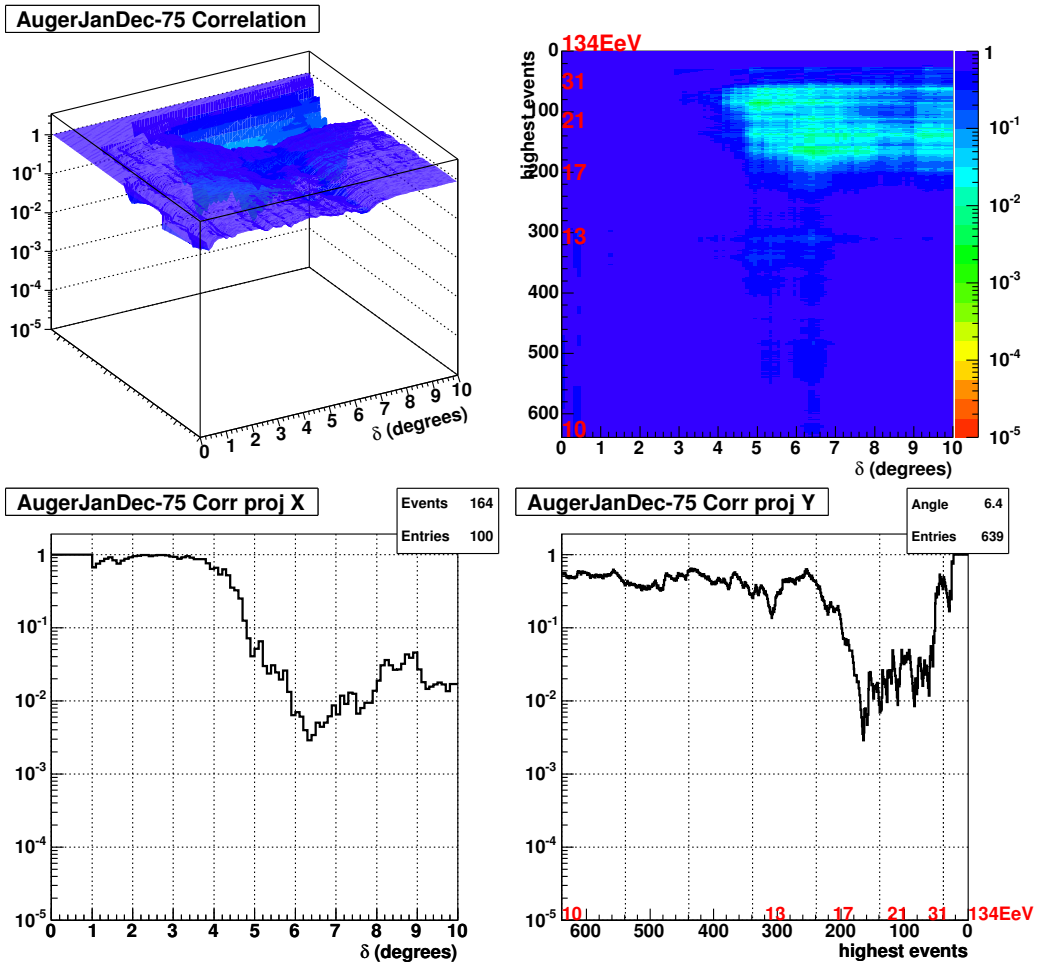


Figure 9.11: Two dimensional scan for significance of auto-correlation of Auger data up to 75°. Maximum significance is at $N_c = 164$ and $\theta_c = 6.4°$.

Top Left: 3-D presentation, x-axis is separation of correlated points in degrees, y-axis is energy threshold, z-axis is probability.

Top Right: top view of 3-D plot, on y-axis, red numbers indicate energy threshold in EeV, color indicates probability.

Bottom Left: Slice of 3-D plot through point of maximum significance projected onto angular separation axis.

Bottom Right: Section of 3-D plot through point of maximum significance projected onto energy threshold axis.

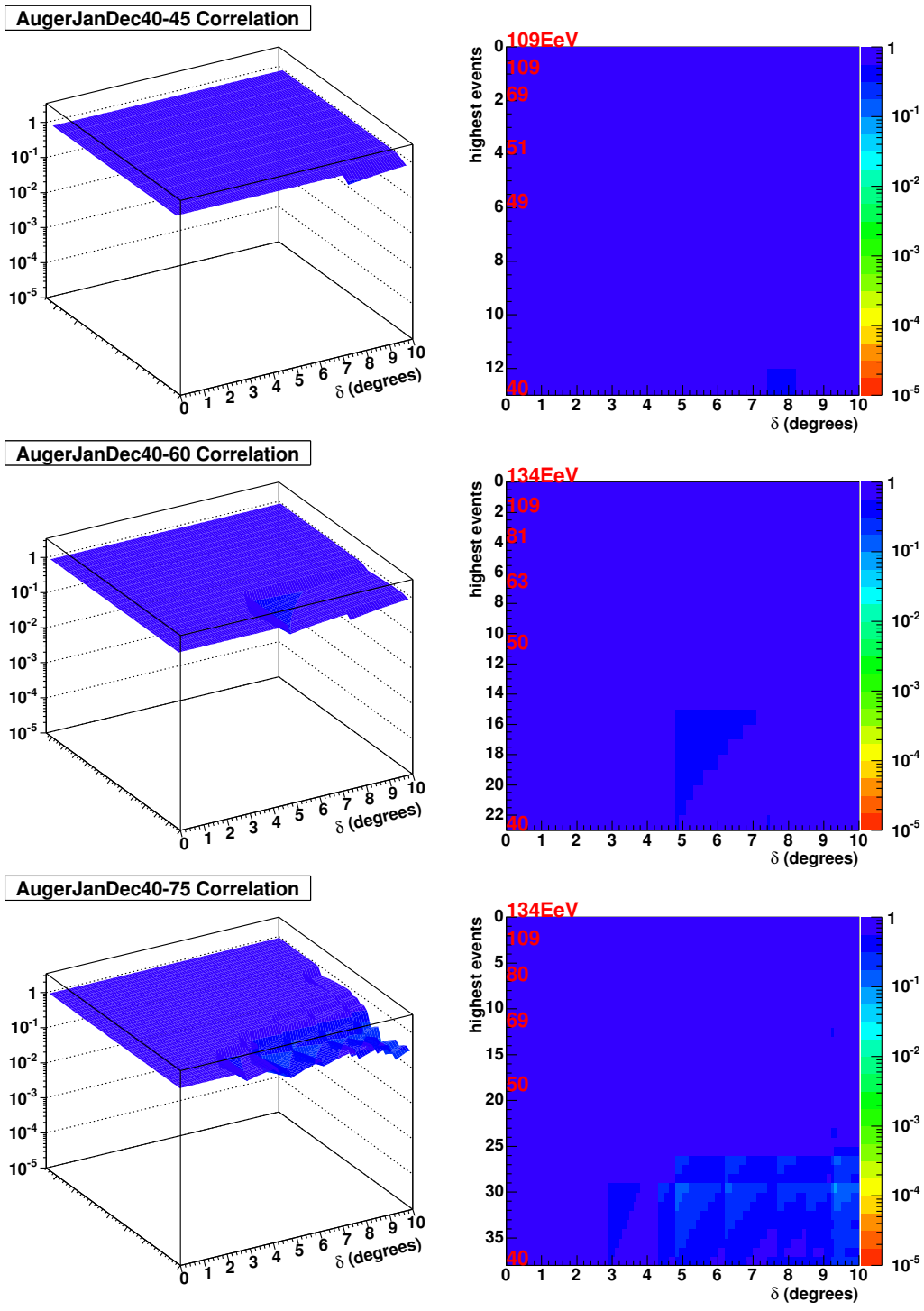


Figure 9.12: Two dimensional scans for significance of auto-correlation of Auger data with energies > 40 EeV for zenith angles of 45° (top), 60° (middle) and 75° (bottom).

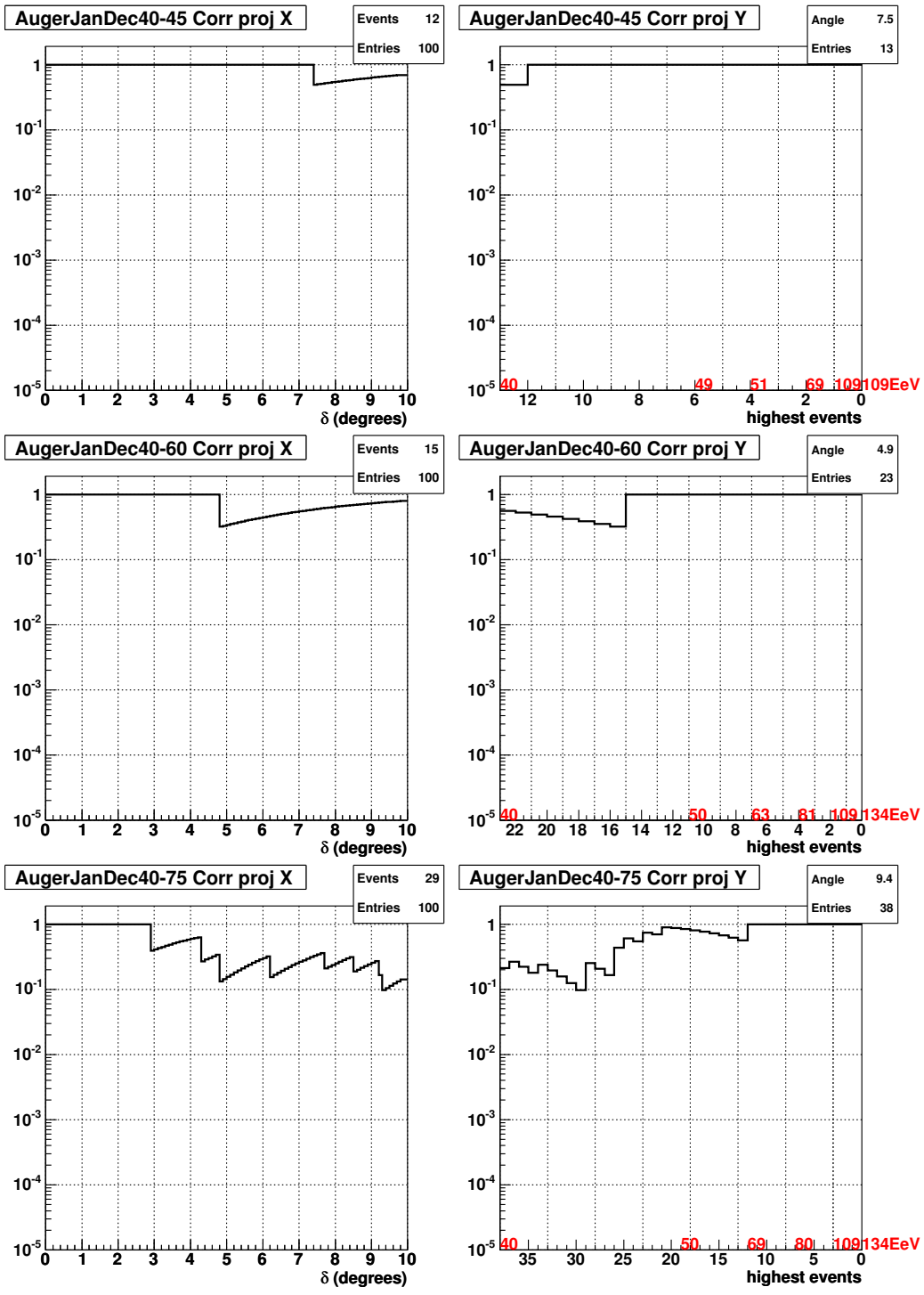


Figure 9.13: X-Y projections of Figure 9.12 for significance of auto-correlation of Auger data with energies > 40 EeV for zenith angles of 45° (top), 60° (middle) and 75° (bottom).

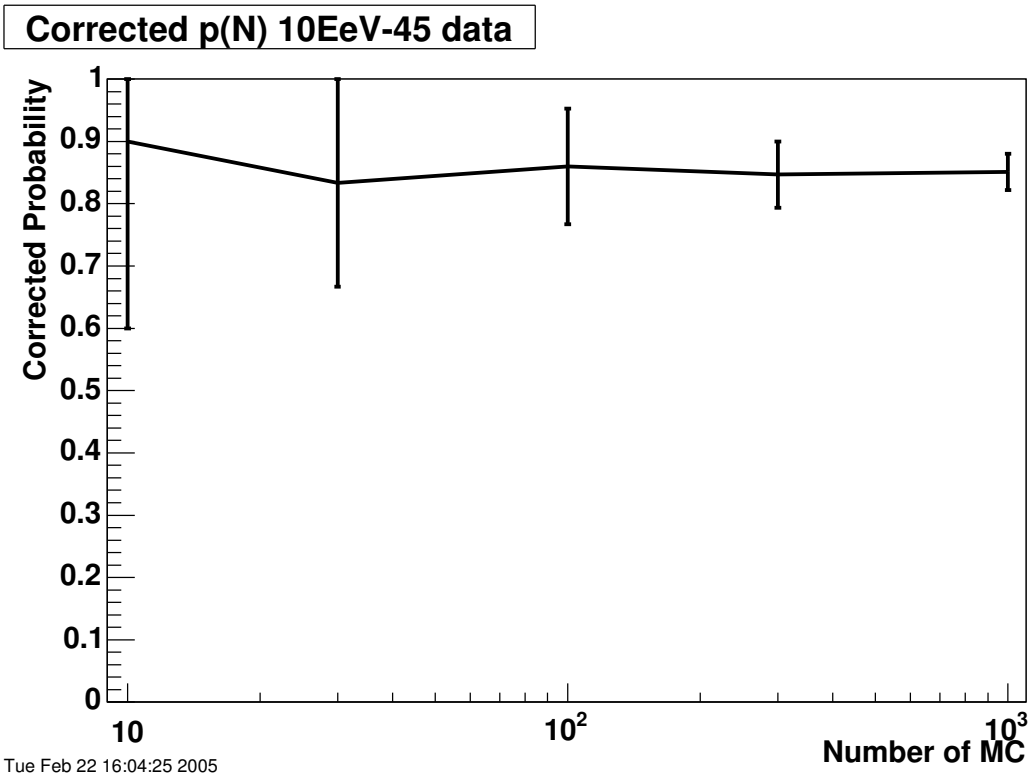


Figure 9.14: Dependence of corrected probability on the number of simulated skies used. Error bar is $sqrtN$

CHAPTER 10

Conclusion

In order to determine if there is clustering in the southern sky, we have developed techniques that reproduce previous studies and obtained no auto-correlation in the first year of Auger data. More specifically: We have reproduced the major results of auto-correlation analysis of AGASA data reported in [22], where they reported P_{min}= P_{data}(N_c=36, θ_c =2.5°)= 8.4×10^{-5}, and we obtained P_{min}= P_{data}(N_c= 35, θ_c =2.4°)= 8×10^{-5}. The agreement is reasonable. We have performed an auto-correlation analysis of HiRES data, and the signal strength of P_{min}∼0.32 is not significant and consistent with published results. In a similar analysis on the first year of Auger data with reconstructed energy in excess of 10 EeV and with zenith $< 75°$ we found 0.003 chance probability in auto-correlation. After correcting for penalty using many random data sets, the chance probability grew to 0.3. In the future as the Auger dataset grows, these tests can be performed again with much higher statistics to find correlations.

References

[1] GE Silicones. Sealants, adhesives, gels, encapsulants: Materials and applications. Technical report, GE Silicones, 2002. available on the web at: http://www.gesilicones.com/silicones/americas/business/workshop/default.shtml.

[2] B. Genolini et al. Design of the photomultiplier bases for the surface detectors of the pierre auger observatory. Technical Report GAP-2001-021, Pierre Auger Observatory, 2001.

[3] B. Genolini et al. Low power high dynamic range photomultiplier bases for the surface detector of the pierre auger observatory. Technical Report GAP-2002-038, Pierre Auger Observatory, 2002.

[4] N. B. Fuqua. *Reliability Engineering for Electronic Design*. Marcel Dekker, Inc., 1987.

[5] P. D. J. Clark and A. Dye. A draft environmental stress screening and burn in test procedure for electronic equipment used within the pierre auger observatory. Technical Report GAP-2002-002, Pierre Auger Observatory, 2002.

[6] A. Tripathi et al. A systematic calibration of surface detectors using muon data from the engineering array. Technical Report GAP-2002-046, Pierre Auger Observatory, 2002.

[7] C. Jillings et al. Results and procedures of testing of pre-production pmts for the surface detector. Technical Report GAP-2002-037, Pierre Auger Observatory, 2002.

[8] P. Allison et al. Surface detector calibration in the engineering array. Technical Report GAP-2002-028, Pierre Auger Observatory, 2002.

[9] A. Lopez-Aguera et al. Signal uncertainty induced by pedestal suppresion procedure. Technical Report GAP-2003-038, Pierre Auger Observatory, 2003.

[10] P. Kurnadi et al. Analysis of the monitoring data from pre-production tanks. Technical Report GAP-2003-061, Pierre Auger Observatory, 2003.

[11] M. Urban et al. Absolute and relative calibration of the surface detectors with the t2 events. Technical Report GAP-2002-012, Pierre Auger Observatory, 2002.

[12] T. McCauley and T. Paul. Geant4 simulation of the surface detectors. Technical Report GAP-2001-018, Pierre Auger Observatory, 2001.

[13] S. Dagoret-Campagne. The sdsim user guide. Technical Report GAP-2002-072, Pierre Auger Observatory, 2002.

[14] K. Arisaka et al. First study and calibration of extended dynamic range pmts and bases with muons using the fermilab water tank. Technical report, Pierre Auger Observatory, 2001.

[15] A. Chou X. Bertou, C. Grunfeld. Preliminary analysis fo the didi preproduction tank. Technical report, Pierre Auger Observatory, 2003.

[16] G. R. Fernandez et al. Surface detector response using lookup table based on geant4 simulation. Technical report, Pierre Auger Observatory, 2004.

[17] G. Rodriguez-Fernandez A. Lopez-Aguerra, V. M. Olmos-Gilbaja. Direct light in inclined showers. Technical report, Pierre Auger Observatory, 2003.

[18] N. Hayashida et al. Updated agasa event list above 4×10^{19}ev. *Astrophys. J.*, 522:225 appendix, 1999. also arXiv: astro-ph/0008102.

[19] N. Hayashida et al. Possible clustering of the most energetic cosmic rays within a limited space angle observed by the akeno giant air shower array. *Phys. Rev. Lett.*, 77:1000, 1996.

[20] M. Takeda et al. Small scale anisotropy of cosmic rays above 1019 ev observed with the akeno giant air shower array. *Astrophys. J.*, 522:225, 1999. also arXiv: astro-ph/9902239.

[21] M. Teshima et al. The arrival direction distribution of extremely high energy cosmic rays observed by agasa. In *Proceedings of the 28th International Cosmic Ray Conference*, pages 437–440, 2003.

[22] C. Finley and S. Westerhoff. On the evidence for clustering in the arrival directions of agasa's ultrahigh energy cosmic rays. *Astroparticle Physics*, 21:359–367, 2004.

[23] R. U. Abbasi et al. Search for point sources of ultra-high energy cosmic rays above 4.0×10^{19}ev using a maximum likelihood ratio test. *accepted for publication in Astrophys. J.*, 2004. also arXiv: astro-ph/0412617.

[24] P. G. Tinyakov and I. I. Tkachev. Correlation function of ultra-high energy cosmic rays favors point sources. *JETP Lett.*, 74:1–5, 2001. also arXiv: astro-ph/0102101.

[25] R. U. Abbasi et al. The high resolution fly's eye collaboration, study of small-scale anisotropy of ultrahigh energy cosmic rays observed in stereo by hires. *Astrophys. J.*, 610:L73, 2004. also arXiv: astro-ph/0404137.

[26] S. Westerhoff et al. The high resolution fly's eye collaboration, search for small-scale anisotropy of cosmic rays above 1019 ev with hires stereo. *Nucl. Phys. B (Proceedings Suppl.)*, 136C:46–51, 2004. also arXiv: astro-ph/0408343.