

**EXTENSION OF THE ICETOP ENERGY SPECTRUM TO 250 TeV
AND APPLICATION OF THE CONSTANT INTENSITY CUT
METHOD TO ICETOP DATA**

by

Ramesh Koirala

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics

Spring 2019

© 2019 Ramesh Koirala
All Rights Reserved

**EXTENSION OF THE ICETOP ENERGY SPECTRUM TO 250 TeV
AND APPLICATION OF THE CONSTANT INTENSITY CUT
METHOD TO ICETOP DATA**

by

Ramesh Koirala

Approved: _____
Edmund R. Nowak, Ph.D.
Chair of the Department of Physics and Astronomy

Approved: _____
John Pelesko, Ph.D.
Interim Dean of the College of Arts and Sciences

Approved: _____
Douglas Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Thomas K. Gaisser, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

David Seckel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Jamie Holder, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Qaisar Shafi, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Richard Braun, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of my advisor Dr. Thomas K. Gaisser. I would like to thank you for your guidance, limitless patience, and enthusiasm to teach during this challenging but fun journey. Your willingness to listen and correct a novice student, especially in the beginning of the journey, enabled me to explore new ideas, and later implement new techniques. For this, I am eternally grateful.

The analyses presented in this dissertation require an in-depth knowledge of the IceTop detector, physics, and computer programming skills. I would like to thank Dr. Serap Tilav for teaching me about the IceTop detector. I would like to thank Dr. Javier Gonzalez and Dr. Hershul Pandya for helping me hone my programming skills. Our regular conversations related to my work and cosmic ray physics in general helped me to progress a lot faster.

Our local group at Bartol Research Institute (Dept. of Physics and Astronomy at the University of Delaware) has always challenged me to think critically and has given valuable comments and suggestions to improve. I would like to thank Dr. Todor Stanev, Dr. David Seckel, Dr. Dennis Soldin, Dr. Pierre-Simon Mangeard, and Dr. Alan Coleman at Bartol. Dr. David Seckel's ability to perform quick calculations and identify microscopic features/differences in plots will always amaze me.

I would like to thank the IceCube Collaboration for creating such an excellent environment for a new researcher like me to grow and flourish. Regular boot-camps and collaboration meetings provided many opportunities to learn and meet many scientists from all over the world. This widened my view on the importance of collaborative effort to achieve a common goal.

I would like to thank Dr. Javier Gonzalez, Dr. Jamie Holder, and Dr. Felix Riehn for introducing me to rock climbing. This sport gave me a reason to be active in our otherwise sedentary lifestyles and helped me to relax. I would like to thank past and current members of Nepali Student Association (NSA) at UD for being happy in my success and uplifting me during periods of low.

I would like to thank all my teachers and mentors, past and present. Especially, I want to thank my high-school teacher from Nepal, Mr. Hukum Prasad Dahal, for accepting me in his school and for his unmatched style and love of teaching. It is during this time I developed my fascination towards physics and mathematics.

I would like to thank my wife Dr. Shree Baba Pokharel. We went through so much together over past decade and I love the way you bring ease and laughter at home. I would like to thank my brother Jiban Koirala and my sister Ramala Koirala (Gadtaula) for your constant love and support. Finally, a very special thank you to my father, Guru Prasad Koirala, and my mother, Chhalu Maya Koirala. I would like to dedicate this work to you. I could not have made it this far without your sacrifice, hard work, infinite encouragement, love, and support.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xiii
ABSTRACT	xxi
 Chapter	
1 INTRODUCTION	1
1.1 Source of Cosmic Rays	3
1.2 Acceleration	7
1.3 Propagation	10
1.4 Composition	12
1.5 Extensive Air Showers	14
1.6 Cascade Equation	16
1.7 Detection Method	17
 2 DETECTOR	 20
2.1 InIce	20
2.2 IceTop	22
2.3 Digital Optical Module (DOM)	25
2.4 Data Transfer	27
2.5 Calibration	28
2.5.1 Calibration of DOM electronics and PMT	28
2.5.2 VEMCal	29
2.5.3 RAPCal	31
 3 NEW TRIGGER AND FILTER	 34
3.1 Background Study	35
3.2 IceTop Volume Trigger	38
3.3 Trigger/Filter Performance	39

4	SIMULATION	41
4.1	GCD File	45
4.2	Level1	46
4.3	Level2	47
4.4	Composition Model	50
4.5	Weighting the Spectrum with the Composition Model	52
4.6	Data-Simulation Comparison	53
4.7	Sybill2.1 Vs QGSJETII-04	54
5	ICETOP AIR SHOWER RECONSTRUCTION	57
5.1	Machine Learning	58
5.1.1	Random Forest Regression	58
5.2	Machine Learning Features	60
5.2.1	Shower Core with Center of Gravity (COG)	61
5.2.2	Direction of Plane Shower Front ($\theta_{\text{plane}}, \phi_{\text{plane}}$)	61
5.2.3	Average Z in Shower Coordinates (ZSC_{avg})	61
5.2.4	Tank Distance from Shower Core (hitR)	62
5.2.5	Other	63
5.3	Reconstruction using Spark	64
5.3.1	Reconstruction of Core Position	65
5.3.2	Reconstruction of Zenith Angle	67
5.4	Reconstruction using Scikit-Learn	68
5.4.1	Reconstruction of Energy	68
5.5	Reconstruction Quality	69
5.6	Data and MC Comparison	71
6	ICETOP LOW ENERGY SPECTRUM	75
6.1	Quality Cuts	76
6.2	Duration	77
6.3	Effective Area	78

6.4	Bayesian Unfolding	79
6.5	Pressure Correction	84
6.6	Systematic Uncertainties	87
6.6.1	Uncertainty due to Composition	87
6.6.2	Uncertainty from Unfolding	89
6.6.3	Uncertainty from Effective Area	90
6.6.4	Uncertainty due to Pressure	90
6.6.5	Uncertainty in VEM Calibration	91
6.6.6	Uncertainty due to Snow	92
6.6.7	Uncertainty due to Duration	93
6.6.8	Total Systematic Uncertainty	94
6.7	Statistical Uncertainties	95
6.8	Cosmic Ray Flux	96
6.9	Tests	98
6.9.1	Different Mean Energy and Initial Prior	99
6.9.2	Different Bin Size	100
6.10	QGSJetII-04 as Hadronic Interaction Model	100
7	CONSTANT INTENSITY CUT METHOD	104
7.1	Motivation	104
7.2	Shower Reconstruction	105
7.3	Data Features	106
7.4	Gaisser-Hillas Type Function	108
7.5	Attenuation Curve for Different Intensity Cuts	110
7.6	Derivation of Cosmic Ray Flux	111
8	RESULTS, DISCUSSION, AND FUTURE WORK	115
8.1	Low Energy Spectrum	115
8.2	Discussion	117
8.3	Constant Intensity Cut (CIC) Method	120
8.4	Future Work	120
	BIBLIOGRAPHY	122
	Appendix	
	A SPECTRUM RELATED	133

B COMPOSITION RELATED 135
C CIC RELATED 136

LIST OF TABLES

3.1	All filter conditions before the addition of two-station trigger and filter. SMT6 for IceTopSMT, STA3 for IceTopSTA3_13, STA5 for IceTopSTA5_13, ILIT for InIceSMT_IceTopCoincidence_13, and InFill3 for IceTop_InFill_STA3_13 is used. IceTopSMT6 trigger is passed on events for first three filters. InIceSMT8 trigger is passed on events for InIceSMT_IceTopCoincidence_13 filter.	35
3.2	Four pairs of nearby infill stations and distance between each pair in meter.	38
4.1	List of CORSIKA datasets that are used in this analysis. Hadronic interaction models used are Sibyll2.1 and QGSJetII-04. Two analyses are performed; one with Sibyll 2.1 and the other with QGSJetII-04 as hadronic interaction models.	42
4.2	Parameters for April Atmosphere used in the simulation. Only first line of Eq. 4.1 is used for all five layers. h is in unit of km and IceTop is located in Layer1.	43
4.3	Resampling radius based on energy bin.	47
4.4	Parameters for H4a composition model.	51
6.1	Correction factor on the final flux due to difference in atmospheric pressure between simulation and 2016 data.	86
6.2	Systematic uncertainty due to iterative Bayesian unfolding.	90
6.3	Systematic uncertainty due to effective area. Uncertainties for each energy bin are calculated with respect to the fit.	90
6.4	Systematic uncertainty due to difference in atmospheric pressure.	91
6.5	Total systematic uncertainty after adding individual systematic uncertainty in quadrature.	95

8.1	Information related to all-particle cosmic ray energy spectrum using two stations events. Sibyll2.1 is the hadronic interaction model assumption. Refer to the text for detail description of each column.	117
A.1	Parameter after fitting effective area with sigmoid type function in Eq 6.6. The first row is <i>Effective area</i> (Aeff) parameters for events with $\cos\theta \geq 0.9$ assuming H4a composition model. The second row is Aeff parameters assuming all Proton composition. The third row is Aeff parameters assuming all Iron composition.	133
A.2	Information related to all-particle cosmic ray energy spectrum using two-station events and assuming Sibyll 2.1 as the hadronic interaction model.	134
A.3	Information related to all-particle cosmic ray energy spectrum using two-station events and assuming QGSJetII-04 as the hadronic interaction model.	134
A.4	Quality of reconstruction. The first row shows the core resolution in meter. The second row shows the zenith resolution in degree. The third row shows the energy resolution. All these resolutions are one sigma of true minus reconstructed value.	134
B.1	Parameters for modified Polygonato composition model. An exponential cutoff in rigidity of population 1 is performed here compared to the original Polygonato composition model.	135
C.1	$\log_{10} S_{125}$ values used in attenuation curve for each curve generated for some constant number of events cut. $\log_{10} S_{125}$ values for the corresponding slant depth (g/cm ²) (X) is tabulated below.	136
C.2	Parameter after fitting attenuation curve with Gaisser-Hillas function. p_1 represents the mean shower maximum X_{\max}	136

LIST OF FIGURES

1.1	Cosmic ray flux measured by various experiments from 10^{12} eV to more than 10^{20} eV and published by the Auger Collaboration [5]. The flux is scaled by $E^{2.5}$ to see features.	3
1.2	An updated version of Hillas diagram from [7]. Initial Hillas diagram is shown in [6]. The diagram shows the boundary criteria for the production of cosmic rays. The solid diagonal lines are the minimum condition to produce protons and irons with energy 10^{20} eV. Possible sources based on their size and strength of the magnetic field are shown.	4
1.3	The difference of ionization at various altitude and that at sea level measured by Kohlhörster in 1913 and 1914. Source of the data is page 7 of [11].	4
1.4	Left: Schematic of the physical situation for first order Fermi acceleration. It occurs in the shock front. The direction of a particle away from the acceleration region, θ_2 , can have values from 0 to $\pi/2$. Right: Schematic of the physical situation for second order Fermi acceleration. It occurs in the plasma cloud. The direction of the particle away from the acceleration region can be any direction with θ_2 ranging from 0 to π	8
1.5	Left: Comparison between elemental abundances in sun and cosmic rays. Data is collected from the book ‘High Energy Astrophysics, Volume 1, 2nd edition’ by Malcolm S. Longair. Right: Boron to Carbon flux ratio of cosmic rays as a function of kinetic energy per nucleon from AMS-02 [61].	11
1.6	Three populations of the modified H4a composition model [62] and their sum. The third population consists of mixed elements with cutoff rigidity 2 EV and only protons with cutoff rigidity 60 EV. . .	14
1.7	Left: Schematic of Heitler’s model for electromagnetic air shower. Right: Schematic of Heitler’s model for hadronic air shower.	15

2.1	Schematic of the IceCube Neutrino Observatory showing IceTop and InIce. IceTop is at the surface of almost 3 km thick ice-sheet. InIce extends from 1450 m to 2450 m below the surface [88].	21
2.2	IceTop geometry showing exact position of all 162 tanks. Two tanks are placed in distance of 10 m and forms a station. Distance between nearby stations is around 125 m. Area surrounded by an irregular polygon at the center has nearby infill stations.	21
2.3	Schematic of an IceTop tank.	23
2.4	A piece of IceTop tank and a US nickel for scale. The black top is the outer polyethylene layer of a tank. The white bottom layer is the layer of reflective zirconium.	24
2.5	Left: Schematic of a DOM. Right: A block diagram of the DOM mainboard. Both figures are from [87]	25
2.6	An example waveform. Channel 0 of ATWD is saturated so channel 1 of lower gain is used for signal digitization. The digitized value of waveform of each <i>Analog Transient Waveform Digitizer</i> (ATWD) bin is called an ADC count.	26
2.7	Experimental muon spectrum on DOM 81-61 fitted with Eq 2.1.	30
2.8	Relative charge difference of low-gain DOM and high-gain DOM of tank 81A.	31
2.9	Diagram of the RAPCal time calibration scheme. Figure is from [88].	32
3.1	Left: The distribution of time difference between hits on DOMs 61 or 62 and 63 of each tank of a station is shown in blue and green. The distribution of time difference between hits on DOM 63 of each station of a nearby infill station pair is shown in red. Right: Six infill stations which form four nearby infill stations pairs are shown along with a circle around each tank of radius 60m.	36
3.2	Left: Cumulative rate with hit on at least one nearby infill station pair. Data used is 2013 FRT from July to December. Right: Histogram of InIce DOMs that are hit along with four high gain DOMs hit in IceTop.	37

3.3	Left and right plots on top show differential station rate from 2016, 2017 and 2018 data. From 2016 data, the average rate of all runs, and from 2017 and 2018 data, rate for test runs is shown. The bottom left plot shows Q_{total} distribution for the 2017 reference run and the 2018 24hr test run. The bottom right plot shows Q_{hit} distribution for the 2017 reference run and the 2018 24hr test run. All four plots show that Trigger/Filter are working as expected.	40
4.1	Difference between the July 2010 South Pole atmosphere, data collected by the <i>Antarctic Meteorological Research Center</i> (AMRC) balloon and the NASA <i>Atmospheric Infrared Sounder</i> (AIRS) satellite, and the July 1997 atmosphere (ATMOD12) used in CORSIKA. The top plots show the temperature and air density difference. There is unphysical behavior seen at 4 km, 10 km, and 40 km. The bottom plot shows the difference in atmospheric depth. Source of these figures is from an IceCube cosmic ray working group internal study performed by Sam DeRidder.	44
4.2	Height of snow accumulated on top of each IceTop stations in 2016. The snow heights used in simulation is from October 2016 measurement, and snow height for data is the average snow height for 2016 data taking period. The difference in the total snow on top of all stations between simulation and data is 1.41 m, which is small given that the total snow height is ~ 142 m spread over an area of 1 km^2	46
4.3	The true core position of simulated showers that pass IceTop volume trigger. Showers are contained within 200 m, 400 m, 600 m, and 800 m depending on their primary energy.	48
4.4	IceTop geometry with nearby InFill area inside the irregular polygon at the center. Re-sampling area for various energy bins is shown by circles. Area increases as energy of the shower increases. Radial distance increases by 200 m for one decade energy increase. Showers are re-sampled within 200 m for energy 10 TeV to 100 TeV, 400 m for energy 100 TeV to 1 PeV and so on.	49
4.5	Left: modified H4a composition model where the individual flux of proton, helium, oxygen, iron, and total is scaled by $E^{2.65}$. Right: Fractional composition of modified H4a model in five elemental groups. CNO+Mg-Si is used in the analysis and is represented by O*.	

4.6	Composition fraction of commonly used 4 different composition models.	52
4.7	Charge histogram of tank 81A compared between data and simulation.	54
4.8	Comparing quantities between Sibyll2.1 and QGSJetII-04. Top: First Interaction height between Sibyll2.1 and QGSJetII-04 for four elements. Bottom: Shower maximum between Sibyll2.1 and QGSJetII-04 for four elements.	56
5.1	Schematic of a decision tree. Data is split into two parts by a feature that minimizes Residual Sum of Squares (RSS) at each node.	59
5.2	Schematic of a plane shower front hitting IceTop tanks. Z_{avg} is the average of z in the shower coordinate system. r is the distance from the shower core to a hit tank.	62
5.3	Feature Importance of 2D features to predict air shower's core position (X and Y) and zenith angle (θ). The maximum feature importance saturates before reaching 35, so it is a good idea to use information from first 35 tanks instead of all 162 tanks.	64
5.4	Reduce in time while running a job in multiple cores using Spark. Time required to complete an example job reduces from 732s using 1 core to 107s using 16 cores.	65
5.5	True vs reconstructed (predicted) x -coordinate and y -coordinate of core positions before final quality cuts.	66
5.6	True vs reconstructed zenith angle before final quality cuts.	68
5.7	True vs reconstructed energy before final quality cuts.	69
5.8	Feature importance to predict x and y coordinate of core position, zenith angle, and energy. List of coordinates of hit tanks has the highest feature importance to predict core position. The zenith angle assuming plane shower front and time of hits have a higher feature importance to predict zenith angle. List of charge on hit tanks and their distance from the shower core have a higher feature importance to predict energy.	70

5.9	Resolution is defined as one standard deviation of true minus reconstructed distribution. Left: Core resolution in meter; Middle: zenith resolution in degree; Right: energy resolution in unit-less quantity. The difference distribution comes from the difference of $\log_{10}[E/\text{GeV}]$ value of true and reconstructed. See Table A.4 for resolution values.	71
5.10	Top: Position of tanks that have been hit. Bottom: COG position of shower core.	72
5.11	Top: Predicted X and Y coordinate of core position of showers. Bottom: Charge on each hit tanks and the distance of hit tanks from the predicted showers' core position.	73
5.12	Left: Comparison between total charge histogram of events from data and simulations. Right: Comparison between total number of stations histogram from data and simulations.	74
6.1	An example fit to the histogram of time difference between consecutive events with an exponential decay function. Parameters of the fit is used to calculate duration on run-by-run basis. The difference in duration provided during data taking and calculated by this method is negligible.	78
6.2	Effective area calculated using MC generated with H4a composition model and Sybill 2.1 hadronic interaction model. A sigmoid function is used to fit the effective area.	80
6.3	Effective area calculated using MC generated with H4a composition model and Sybill 2.1 hadronic interaction model. A sigmoid function is used to fit the effective area. Parameters of all fits are given in Appendix A. Left: Effective area for two zenith bins. Right: Effective area assuming all proton, all iron, and four elements composition.	80
6.4	An example plot showing an evolution of systematic uncertainty in each iteration for iterative Bayesian unfolding in linear (left) and log (right) y-axis. Iter1 is the bottom line and iter100 is the top line. The mathematical formulation of uncertainty and its propagation in each iteration are given in [123]. The method is implemented in pyUnfolding package.	81

6.5	Response Matrix calculated from simulation. An element of a response matrix is a fraction of events in true energy bin distributed over the reconstructed energy bin. In Bayes theorem in Eq 6.7, $P(E C)$ represents a response matrix.	82
6.6	Top: Unfolding matrix for the first iteration (left) and the last iteration (right). Bottom: Energy distribution after each iteration of Bayesian iterative unfolding. Iteration was done until termination criteria were met.	84
6.7	Left: The rate of two-station events on run-by-run basis compared with the average South Pole atmospheric pressure during data-taking period. Right: Pressure histogram for 2016 South Pole atmosphere with a mean of $\sim 691 \text{ g/cm}^2$	85
6.8	Percentage deviation of cosmic rays flux when atmospheric pressure is $\sim 698 \text{ g/cm}^2$ from the flux when pressure is $\sim 691 \text{ g/cm}^2$. This deviation is used as the correction factor to correct the final flux. The error on the correction factor is used as the systematic uncertainty due to Pressure difference between average pressure of 2016 South Pole atmosphere and pressure due to atmosphere profile used in simulation.	86
6.9	Top: Percentage spread of cosmic ray flux assuming different composition models. Maximum spread is used as the systematic uncertainty due to composition. Bottom: Systematic uncertainty due to composition relative to the H4a composition model.	88
6.10	Systematic uncertainty for each of twelve iterations in iterative Bayesian unfolding. The twelfth systematic uncertainty is used as the systematic uncertainty from unfolding procedure. The mathematical formulation of uncertainty and its propagation in each iteration is shown in [123]. The method is implemented in pyUnfolding package.	89
6.11	Effective area calculated using MC generated with Sibyll2.1 hadronic interaction model and using H4a as composition model. Solid black line is the fit with Equation 6.6 and a gray band is an error of the fit. Energy region with efficiency higher than 50% is used in this analysis.	91
6.12	Left: Histogram of difference between mean value of VEMCal for 2016 in the unit of pe and value used in simulation. The difference is calculated on DOM-by-DOM basis. Right: Distribution of deviation of 1 VEM in pe for 2016 from its mean on DOM-by-DOM basis.	92

6.13	Snow height on top of IceTop geometry. Snow height on top of each station is shown in Fig 4.2. To compare snow heights between data and simulation, an average value of 2016 is used for data and October 2016 snow height is used for simulation.	93
6.14	The plot shows the individual systematic uncertainties for each energy bins. Total systematic uncertainty is the sum of individual uncertainty in quadrature.	94
6.15	Statistical error calculated using initial energy distribution assuming it follows a Poisson distribution and an output after Bayesian iterative unfolding. The statistical error calculated using the pyUnfolding package is used as the final statistical uncertainty which is higher but still small compared to the flux.	95
6.16	Top: Cosmic Ray Flux using IceTop 2016 Data. Bottom: Cosmic Ray Flux using IceTop 2016 Data scaled by $E^{1.65}$	97
6.17	Cosmic Ray Flux using IceTop 2016 Data compared with other experiments. Special thanks to Stephane Coutu for providing with the fit parameters for CREAM energy spectrum.	98
6.18	Left: Effect of different choice of bin center on the cosmic ray flux. Right: Effect of different priors on the cosmic ray flux	99
6.19	All-particle cosmic ray flux for two energy bins to check the effect of different bin size. Iterative Bayesian unfolding accounts for all energy bin migration proving the choice of bin size is arbitrary if there is enough data.	101
6.20	Top: Scaled cosmic ray flux assuming QGSJetII-04 hadronic interaction model and compared with 3-year IceTop energy spectrum. The gray band around the flux is the total systematic uncertainty. Bottom: Comparison of the final flux assuming Sibyll2.1 and QGSJetII-04.	102
7.1	Differential (left) and Integral (right) S_{125} vs $\sin^2\theta$ spectrum. Color represents the number of events in each bin. The integral distribution at a given S_{125} has all events with S_{125} equal to or greater than that value.	107

7.2	Differential (left) and integral (right) S_{125} distribution for each $\sin^2\theta$ bins. The integral distribution at a given S_{125} has all events with S_{125} equal to or greater than that value.	108
7.3	Zenith distribution of 2011 and 2012 data.	109
7.4	Top: Attenuation curves from integral spectra fitted with the Gaisser-Hillas function from Eq 7.7. The highest curve corresponds to the highest energy and the lowest curve corresponds to the lowest energy. Bottom: X_{\max} vs. Energy. X_{\max} is obtained after fitting attenuation curve with Eq 7.7.	113
7.5	Differential spectrum of S_{125} calculated at 27° after correcting atmospheric attenuation for different $\sin^2\theta$ bins. Spectra below 11 VEM start bending away from the trend. Hence, 11 VEM is used as the threshold S_{125} value.	114
7.6	Compare the number of events in higher energy bins between energy spectrum using CIC method (left) and the 3-year IceTop energy spectrum (right). The number of events using 2 years of IceTop data and CIC method is more than the number of events using 3 years of IceTop data.	114
8.1	All particle cosmic ray flux showing the overlap region from 3-year IceTop analysis and this analysis. The energy spectrum from this analysis lies within the systematic uncertainty of 3-year IceTop energy spectrum.	118
8.2	All-particle cosmic ray flux showing Peters cycle based on H4a composition model.	119
A.1	Parameters of South pole atmosphere for Jul. 01, 1997 (MSIS-90-E) after fitting with Eq. 4.1. This table is copied from CORSIKA user-guide.	133

ABSTRACT

The all-particle cosmic ray energy spectrum is an important tool to study acceleration and propagation mechanisms of cosmic rays. The main goal of this dissertation is to lower the energy threshold of the all-particle cosmic ray energy spectrum to overlap with the energy spectra from direct measurements. Previous studies with IceTop/IceCube have measured the spectrum from PeV to EeV. Those measurements used events with at least five-stations hit. In this study, events with two or more stations are used to measure the energy spectrum from 250 TeV to 10 PeV.

In addition to lowering the threshold of the cosmic-ray energy spectrum, this thesis also includes a preliminary application of the constant intensity cut method to IceTop data. This method has the potential to improve the reach of IceTop in the EeV range.

To lower the energy threshold of IceTop, a new trigger and filter is developed and is implemented since May 20, 2016. The new trigger and filter uses 4 pairs of nearby infill stations where the separation between stations in each pair is less than 50 m. Data from the entire array are collected for events in which at least one infill pair is hit. Data to calculate the energy spectrum are collected from May 2016 to April 2017 with a livetime (duration) of 330.43 days. A total of 7,442,086 events is used after all quality cuts.

The constant intensity cut (CIC) method is a widely used method to calculate the cosmic ray energy spectrum. The method increases statistics in energy spectrum analysis by including air showers from higher zenith angle. In this dissertation, CIC is used in particular to study atmospheric attenuation and to measure indirectly the energy-dependent average shower maximum ($\langle X_{max} \rangle$). The use of CIC method to

measure the all-particle cosmic ray spectrum with IceTop in the range of PeV to EeV is discussed.

This dissertation is divided into eight chapters. Chapter 1 is the introduction. Chapter 2 details the IceCube Neutrino Observatory. It describes the Digital Optical Module (DOM) and discusses how DAQ system, trigger/filter, and data transfer operate. IceTop, a surface component of the IceCube Neutrino Observatory, uses various calibration techniques. These calibrations are also discussed in this chapter. Chapter 3 describes the new two-station trigger and filter implemented to collect low energy cosmic rays. Chapter 4 discusses the simulation used for low energy spectrum analysis. Chapter 5 describes the reconstruction of air showers collected using two-station trigger/filter based on machine learning. Chapter 6 reports the result of all-particle cosmic ray spectrum using two-station events. The energy threshold using two-station events is lowered down to ~ 250 TeV. Chapter 7 details the constant intensity cut (CIC) method. This method is studied to increase the zenith range for energy spectrum and to use a zenith independent energy proxy for all-particle energy spectrum calculations. The final chapter, Chapter 8, summarizes the results and discusses possible future work.

Chapter 1

INTRODUCTION

Cosmic rays are charged particles that reach Earth from space. They arrive at a rate of approximately 1000 particles per square meter per second [1] and are almost isotropic. Most cosmic rays are protons and alpha particles. Besides protons and alpha particles, there are small percentages of heavier nuclei, electrons, positrons, and anti-protons. Most cosmic rays have energies in the *giga-electronvolt* (GeV) range and a few have energies as high as 10^{20} *electron-volt* (eV). There are many energetic particles that bombard us; however, there are still open questions about where and how they are produced. Charged particles diffuse while propagating from source to the Earth due to an interstellar magnetic field. They bend from a straight trajectory by the Lorentz force and appear to come from random directions. Therefore, the origin of a very high energy cosmic ray is not accurately known to this date as it is difficult to point back to its source. Additionally, the acceleration and propagation mechanism is not fully understood. To understand more about the source, acceleration, and propagation of cosmic rays, a precise measurement of their energy spectrum and composition is necessary.

The energy spectrum is a distribution of cosmic rays based on their energies. The energy spectrum after taking the exposure of a detector into account is called flux¹. The flux is a detector-independent quantity that can be compared among many experiments. The cosmic ray flux approximately follows a power law. The

¹ In this dissertation, ‘flux’ is used interchangeably with ‘energy spectrum’.

flux decreases sharply with increasing energy. The flux of all-particle cosmic rays can be approximately described by

$$\frac{dN}{dE} \propto E^{-\gamma} \quad (1.1)$$

where dN/dE is the number of cosmic rays with energy in between E and $E + dE$ and γ is the differential spectral index. The value of γ fluctuates around 3. A change in γ produces features in the flux. Around 3×10^{15} eV, γ increases from ~ 2.7 to ~ 3.0 and creates a knee-like structure, first mentioned in 1958 by Kulikov and Khristiansen [2]. Similarly, around 10^{18} eV, γ decreases from ~ 3.0 to ~ 2.7 and creates an ankle-like structure. The knee is believed to be the beginning of the end of cosmic rays from galactic sources [3]. The ankle is believed to be the beginning of cosmic rays from extra-galactic sources [4]. The steepening of the spectrum at the knee is often assumed to depend on the cut-off of elemental flux based on magnetic rigidity. For a fixed rigidity, the energy of a particle is proportional to its charge. The proton flux starts bending at the knee. The helium flux starts bending at twice the maximum energy of a proton. The iron flux starts bending at energy 26 times the maximum energy of the proton. Fig 1.1 from [5] shows the all-particle cosmic ray flux from both ground-based and balloon or satellite-based experiments. The knee around 3×10^{15} eV and ankle around 1×10^{18} eV are visible in the flux.

Not all celestial objects can produce high energy cosmic rays. The size of the source must be greater than twice the Larmor radius to contain and accelerate charged particles. The accelerator must fulfill the Hillas criteria [6], i.e., $BL > 2E/Z\beta$ where B is the magnetic field strength in μG , L is the size of accelerator in *parsec* (pc), Z is the atomic charge of a particle, βc is a characteristic velocity of scattering center, and E is the energy in PeV. Fig. 1.2 shows the plot of the magnetic field strength of potential accelerators vs. size. The solid diagonal lines are boundary conditions for a source to produce protons and irons of energy 10^{20} eV. The maximum energy a source can produce is given by $E_{max} = \beta Z e B R$, where R is the size of the accelerator.

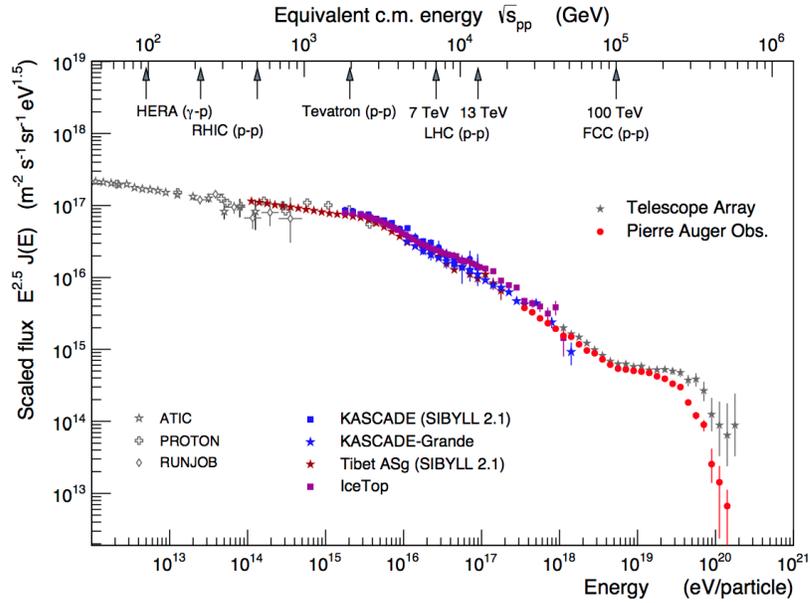


Figure 1.1: Cosmic ray flux measured by various experiments from 10^{12} eV to more than 10^{20} eV and published by the Auger Collaboration [5]. The flux is scaled by $E^{2.5}$ to see features.

1.1 Source of Cosmic Rays

In the early 1890s, the source of radiation detected on Earth was unknown to physicists. In the quest to solve this puzzle, Victor Hess went 5.2 km up in the atmosphere in a balloon in 1912. He reported that the radiation increases with altitude after an initial reduction [8]. He concluded that the source of radiation detected on Earth is of cosmic origin. In 1913, Kohlhörster verified Victor Hess' claim by going up another 3.8 km to 9 km [9]. Kohlhörster's measurement of ionization at different altitudes compared to that at the sea level is shown in Fig. 1.3. Erich Regener and Georg Pfozter later studied the phenomenon of increase in radiation with altitude. In 1935, they showed that the number of particles increases up to an altitude where atmospheric pressure is 100 mmHg (~ 17 km) and decreases rapidly after that point at higher altitudes [10]. This study gave a glimpse of a shower profile in the atmosphere.

Kohlhörster's verification of Hess' discovery that high energy cosmic rays continuously bombard the Earth raised a new question of where exactly they are coming

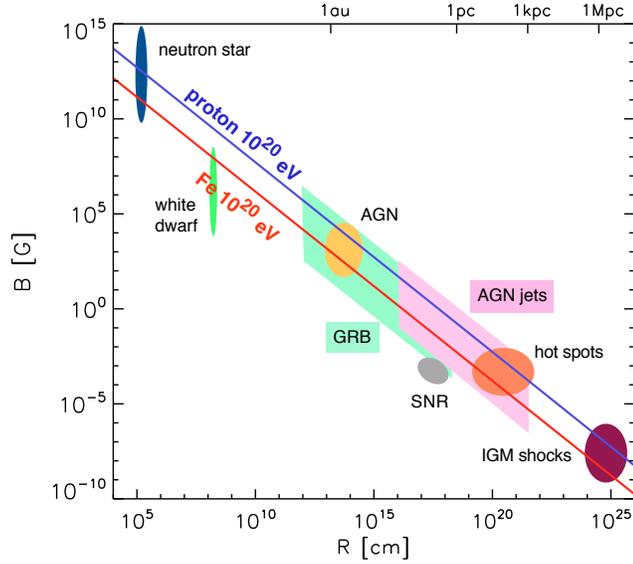


Figure 1.2: An updated version of Hillas diagram from [7]. Initial Hillas diagram is shown in [6]. The diagram shows the boundary criteria for the production of cosmic rays. The solid diagonal lines are the minimum condition to produce protons and irons with energy 10^{20} eV. Possible sources based on their size and strength of the magnetic field are shown.

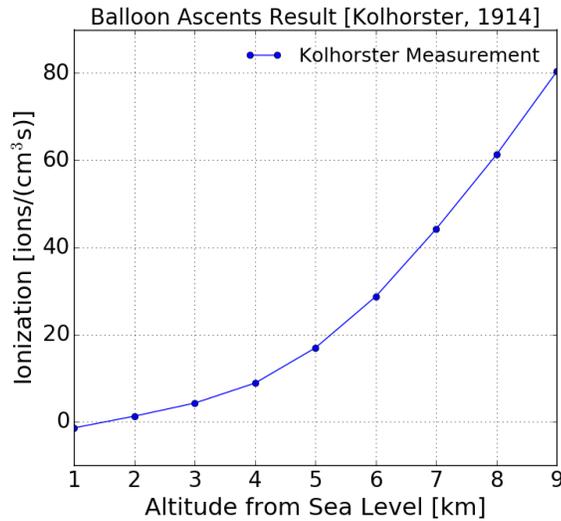


Figure 1.3: The difference of ionization at various altitude and that at sea level measured by Kohlhörster in 1913 and 1914. Source of the data is page 7 of [11].

from. Several telescopes and detectors have been built all around the world to answer this question.

As a result of many years of data collection and analysis, several galactic and extra-galactic candidates have been cataloged as the possible sources of high energy cosmic rays. The possible galactic sources are supernovae, neutron stars, and pulsars. The possible extra-galactic sources are *Gamma Ray Bursts* (GRBs), starburst galaxy, and *Active Galactic Nuclei* (AGN). These sources have an extreme environment that might have accelerated particles to such high energies.

Supernova remnants are possible galactic sources of cosmic rays, with suitable conditions for particle acceleration. When a massive star's fuel is used up, the star collapses and the energy release causes the explosion of the star's outer material, while the inner part collapses to a compact remnant. This explosion generates a shock wave which can accelerate particles by the Fermi mechanism, which is discussed in the next section. The explosion is called a supernova explosion. Supernovae as possible cosmic ray sources were first proposed by W. Baade and F. Zwicky in [12]. Measurements using X-ray telescopes such as the 'Advanced Satellite for Cosmology and Astrophysics' (ASCA) [13] have shown that electrons are accelerated to energy ≥ 200 TeV in supernova remnant shocks [14]. Protons accelerated by a supernova remnant to TeV energies has also been observed [15]. It has been theorized that young supernovae (few months to years) can accelerate cosmic rays to PeV energies [16, 17, 18]. Supernova remnants where cosmic rays are accelerated to high energy, also emit high energy gamma rays and neutrinos. The relation of energy between cosmic ray, gamma ray and neutrino is approximately given by $E_\nu \approx E_\gamma/2 \approx E_N/20$, indicating that, for example, a detection of 10 TeV gamma ray corresponds to approximately 100 TeV cosmic ray. The 'High Energy Stereoscopic System' (H.E.S.S.) [19] collaboration found emission with energy < 1 TeV from nine young supernovae [20]. The 'Major Atmospheric Gamma Imaging Cherenkov' (MAGIC) Telescope [21] and the 'Very Energetic Radiation Imaging Telescope Array System' (VERITAS) [22] collaborations found gamma rays up to 10 TeV from supernova remnant Cassiopeia A [23, 24, 25].

However no direct evidence of PeV cosmic rays from supernovae remnants have been found yet.

Neutron stars are another possible galactic sources that can produce high energy cosmic rays. It is a compact star formed after the supernova explosion of a massive star that failed to form a black hole. It has a strong magnetic field and a radius of about 20 km. A vibrating neutron star [26] or neutron star in a binary system [27, 28, 29] can accelerate particles to very high energies. Another galactic source candidate is a *Pulsating Radio Star* (pulsar). This is a highly magnetized neutron star that emits a beam of electromagnetic radiation, especially radio waves. Low frequency electromagnetic radiation produced in pulsars can accelerate cosmic ray particles to EeV energy range [30]. A detailed description of the neutron star, especially pulsar, as the source of cosmic rays is found in [31, 32].

A potential extra-galactic source is a GRB [33, 34, 35]. The rate of GRB explosions and the energy produced can explain the measured cosmic rays energy spectrum, especially in the ultrahigh energy region (\geq EeV) [36]. GRBs are very energetic, but transient; typically lasting only seconds. GRBs are produced by the collapse of very massive progenitor star [37], or by merging two neutron stars or a binary of a black hole and a neutron star [38, 39].

Another extra-galactic source candidate are starburst galaxies. The rate of star formation, as well as supernova explosions, is higher (approximately a factor of 10 to 100) in starburst galaxies than in average galaxies, as discussed in [40]. A starburst galaxy can be formed when two nearby galaxies encounter each other and interact gravitationally. Sometimes they collide with each other, pushing gas from outer regions towards the center. This causes the density of dust around the center of starburst galaxy to increase. The higher density of dust in the central region of the galaxy triggers massive star formation. These massive stars burn their gas quickly, resulting in supernova explosions at higher rate. Various starburst galaxies as the possible source of high energy gamma rays and cosmic rays have been studied, and some recent work can be found in [41, 42]

Another source candidate is Active Galactic Nuclei (AGN). These are galaxies with a super-massive black hole at the center and a torus of dust and a jet of particles away from the torus plane. Active galaxies are further divided into radio galaxies, quasars, and blazars, based on the observer's line of sight. If an observer is observing an AGN edge-on, a radio source is detected because of the dust particles of the torus blocking other radiations. This AGN is called a radio galaxy, or a Seyfert galaxy. An AGN viewed at an angle is called a quasar. If an observer is observing an AGN with a jet pointed directly towards the observer, it is called a blazar. An AGN can accelerate cosmic rays to the ultrahigh energy region in an expanding shock [43], or in relativistic jets [44], which results in a power law energy spectrum. A description of AGN as the possible source of ultra high energy cosmic rays can be found in [45, 46, 47].

The IceCube Neutrino Observatory detected a high energy neutrino on 22 September 2017 that possibly came from a blazar called TXS 0506+056 [48]. The position of TXS 0506+056 is 77.3582° right ascension and $+5.69314^{\circ}$ declination [49]. Follow-up analysis with data collected during 15 to 27 September 2017 by the 'Large Area Telescope' (LAT) in Fermi Gamma-ray Space Telescope [50] was performed. An increase in the flux of gamma rays by a factor of ~ 6 coming from the same direction was measured [51]. Similar follow-up analysis was performed using 12 hours of data from 28 September 2017 to 03 October 2017 collected by the MAGIC Telescope and measured gamma rays higher than 100 GeV coming from the same direction [52]. Neutrinos and gamma-rays can be produced by collision of cosmic rays with dust near the source. These high energy neutrino events in coincident in direction and time with gamma-ray events from the same source indicate a detection of a possible source of high energy cosmic rays for the first time.

1.2 Acceleration

The energy density (integral) of cosmic rays in the galaxy is about $0.83\text{-}1.02\text{ eV/cm}^3$ [53] and the particle density is roughly 1 proton/cm^3 . The volume of the Milky Way galaxy is $4 \times 10^{66}\text{ cm}^3$ if we assume its shape as a cylinder with radius

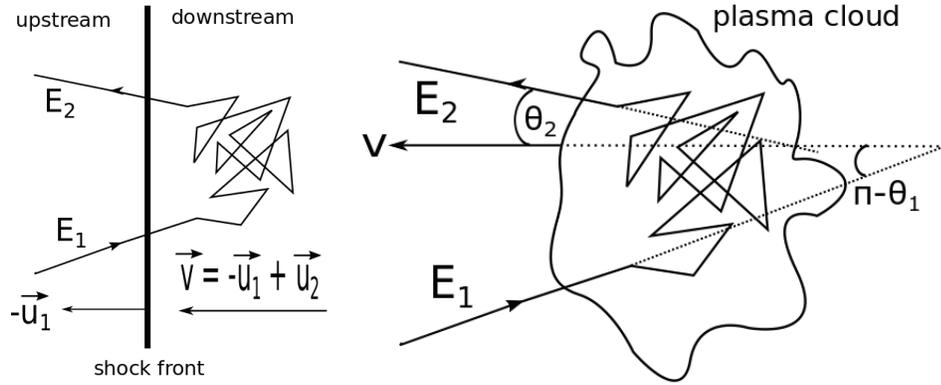


Figure 1.4: Left: Schematic of the physical situation for first order Fermi acceleration. It occurs in the shock front. The direction of a particle away from the acceleration region, θ_2 , can have values from 0 to $\pi/2$. Right: Schematic of the physical situation for second order Fermi acceleration. It occurs in the plasma cloud. The direction of the particle away from the acceleration region can be any direction with θ_2 ranging from 0 to π .

15 kpc and height of 200 pc. The power required to generate cosmic rays, assuming all are galactic, with a residence time of 6×10^6 years is approximately 5×10^{40} erg/sec. The power of 10 solar mass ejecta from a supernova explosion every 30 years with a speed of 5×10^8 cm/s is 3×10^{42} erg/s. Comparing the power generated by a supernova and the power required to produce the observed cosmic rays, it can be said that a few percent of the energy from a supernova is sufficient to energize all galactic cosmic rays. The mechanism that energizes cosmic rays to such a high energy is still being studied.

Many acceleration mechanisms are suggested that can accelerate particles up to more than 10^{20} eV. A method proposed in [54] suggests that in a young supernova remnant a particle can have energy as high as $\sim (10^{17} \times Z)$ eV but not higher than 3×10^{18} eV, where Z is the atomic number of the particle. In the espresso mechanism [55], a powerful AGN can accelerate particles up to energy more than 10^{20} eV if a seed particle has energy around 10^{17} eV. Enrico Fermi proposed a simpler acceleration mechanism in 1949 where charged particles are accelerated by interacting with ‘wandering magnetic field’ in interstellar space [56]. Each time a particle crosses the acceleration region, its energy is changed by an amount proportional to the energy

of a particle with each encounter. This mechanism is called the Fermi mechanism.

The Fermi mechanism is classified into first order and second order based on the type of acceleration region, as shown in Fig 1.4. First order Fermi acceleration occurs at a shock front where the direction of a particle away from the acceleration region is limited. An escape angle, θ_2 , measured from the direction of the motion of an acceleration region can have value from 0 to $\pi/2$. Second order Fermi acceleration occurs in a plasma cloud. The direction of the particle away from the acceleration region can be any direction, with θ_2 ranging from 0 to π .

In the Fermi mechanism, charged particles can have multiple collision-less scattering in a turbulent magnetic field. If there are n encounters with change in energy $\Delta E = \xi E$ per encounter, the final energy of a particle will be $E_n = E_0(1 + \xi)^n$. A charged particle enters the acceleration region and comes out gaining energy by a factor ξ . The factor (ξ) by which a particle gains energy is given by

$$\xi = \frac{1 + \frac{4}{3}\beta + \frac{4}{9}\beta^2}{1 - \beta^2} - 1 \sim \frac{4}{3}\beta \text{ (1st order)} \quad (1.2)$$

$$\xi = \frac{1 + \frac{1}{3}\beta^2}{1 - \beta^2} - 1 \sim \frac{4}{3}\beta^2 \text{ (2nd order)} \quad (1.3)$$

where $\beta = V/c$ and V is the velocity of plasma in downstream with respect to the plasma in upstream for the first order, and the velocity of plasma cloud for the second order Fermi acceleration mechanism. Equations 1.2 and 1.3 are the energy gain factor in each cycle in first order and second order Fermi acceleration respectively. A charged particle gains a larger amount of energy in each cycle by the first order Fermi mechanism compared to the second order.

The Fermi mechanism generates an energy spectrum that follows an inverse power law [56]. The number of particles, N , with energy equal to or greater than E produced by the Fermi mechanism is given by

$$N(\geq E) \propto \left(\frac{E}{E_0}\right)^{-\gamma} \quad (1.4)$$

where E_0 is the initial energy and γ is the spectral index. For first order Fermi acceleration, the spectral index is given by

$$\gamma = \frac{3}{u_1/u_2 - 1} \quad (1.5)$$

where $-\vec{u}_1$ is the velocity of the shock front and \vec{u}_2 is the velocity of the plasma in downstream with respect to the shock front, as shown on the left-hand side plot of Fig 1.4. Similarly, the spectral index from second order Fermi acceleration is

$$\gamma = \frac{1}{\xi c \rho_c \sigma_c T_{esc}} \quad (1.6)$$

where ρ_c is the density of plasma clouds, σ_c the cross section, and T_{esc} is the time to escape the acceleration region.

The maximum energy of a particle which can be reached with the Fermi mechanism in a supernova remnant is estimated to be about 100 TeV. A particle should stay longer in the acceleration region, or the magnetic field of the system should be stronger, for a particle to have higher than 100 TeV energy with this mechanism.

1.3 Propagation

The Milky Way is a spiral galaxy with an estimated 3-5 μG [57] of magnetic field strength roughly parallel to the spiral arms. An accelerated charged particle propagates in space with a magnetic field. The charged particle follows a helical motion in a constant and uniform magnetic field with a net propagation in the direction of the magnetic field and follows a random path in a turbulent magnetic field. The transport equation for these particles is explained in detail in [1] and is given by

$$\frac{\partial N_i(E, x)}{\partial t} + V \cdot \nabla N_i(E, x) - \nabla \cdot (D_i \nabla N_i(E, x)) = Q_i(E, t) - p_i N_i + \frac{\nu \rho(x)}{m_p} \sum_{k \geq i} \int \frac{d\sigma_{i,k}(E, E')}{dE} N_k(E', x) dE' \quad (1.7)$$

where the left side is the term for conservation of particle i and the right hand side is the contribution to the production and decay of i due to various processes. $N_i(E, x)$ is the time dependent density of particles of type i at position x . V is the velocity field,

D is the diffusion coefficient, Q is the rate of particle produced by the source, and p_i is the rate of loss due to collision and decay. The last term on the right side is the cascade term which accounts for the production of particle i from particles with higher atomic number than i and with higher energy.

An approximation is used to solve Eq. 1.7. Commonly used approximations are Leaky box model [58], Nested leaky box model [59], and Diffusion model [60]. The Leaky box model assumes that a particle is trapped in the galactic volume with a constant probability per unit time to escape, but it can propagate freely within that volume. The Nested Leaky box model assumes smaller confinement around the source and a particle leaks from the confinement region in an energy-dependent time. The diffusion model is a more realistic approach where a particle interaction with the interstellar magnetic field is taken into account during propagation, so the diffusion coefficient is not assumed to be a constant. For the scope of this dissertation, only the Leaky box model is discussed as it provides a good description of physics during the propagation of cosmic rays. The detailed explanation of other models can be found in [1].

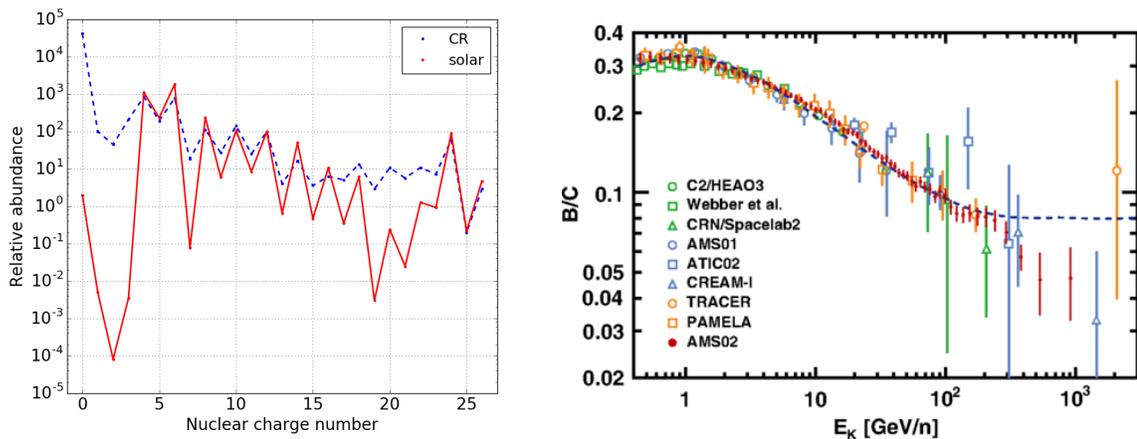


Figure 1.5: Left: Comparison between elemental abundances in sun and cosmic rays. Data is collected from the book ‘High Energy Astrophysics, Volume 1, 2nd edition’ by Malcolm S. Longair. Right: Boron to Carbon flux ratio of cosmic rays as a function of kinetic energy per nucleon from AMS-02 [61].

In the Leaky box model, $V \cdot \nabla N_i(E, x) - \nabla \cdot (D_i \nabla N_i(E, x))$ in Eq. 1.7 is replaced by N/τ_{esc} where τ_{esc} is the time a particle takes to escape. It is assumed to be significantly greater than the time it takes for light to travel the thickness of a galactic disk. This assumption simplifies Eq 1.7 in equilibrium when $N_i(E, x)$ is time independent and converts to

$$\frac{N_i(E)}{\tau_{esc}(E)} = Q_i(E) - \left[\frac{\beta c \rho}{\lambda_i} + \frac{1}{\gamma \tau_i} \right] N_i(E) + \frac{\beta c \rho}{m_p} \sum_{k \geq i} \sigma_{i,k} N_k(E) \quad (1.8)$$

where $\gamma \tau_i$ is the Lorentz dilated lifetime of a particle i and $\lambda_i = \rho \beta c \tau_i$. If Eq 1.8 is implemented to get the ratio of Boron to Carbon density (B/C) measured on Earth, it is further assumed that Boron is produced only from the collision of Carbon and Oxygen with gas in the interstellar medium ($Q_B(E) = 0$). This assumption is made by comparing abundance of Boron in cosmic rays and from the Sun. The higher abundance of Boron in cosmic rays as shown in the left plot of Fig 1.5 is believed to be caused by spallation of higher mass elements during propagation. Additionally, it is assumed that Boron does not decay ($\tau_B = \infty$). Then Eq 1.8 for Boron becomes

$$N_B \left(1 + \frac{\lambda_{esc}}{\lambda_B} \right) = \lambda_{esc}(E) \left(\frac{\sigma_{C \rightarrow B} N_C + \sigma_{O \rightarrow B} N_O}{m_p} \right) \quad (1.9)$$

where $\lambda_{esc} = \beta c \rho \tau_{esc}$. Since N_C and N_O is approximately equal, the ratio of the density of Boron and Carbon is given by

$$\frac{N_B}{N_C} \approx \frac{\lambda_{esc}(E)}{1 + \lambda_{esc}(E)/\lambda_B} \frac{\sigma_{C \rightarrow B} + \sigma_{O \rightarrow B}}{m_p} \quad (1.10)$$

The ratio of the density of Boron and Carbon measured by AMS-02 is shown on the right plot of Fig 1.5, and it can be fitted with an energy-dependent power-law, $E^{-\delta}$, to know the contribution of spallation during propagation on the total cosmic ray flux detected on Earth. The differential spectral index for source spectrum is given by $\gamma - \delta$ where γ is the differential spectral index from Eq 1.1.

1.4 Composition

A particle spends around 6 million years on average, less if energy is higher, in the galaxy before reaching the Earth. A particle reaching Earth can be any element

from the periodic table. Its energy depends on its rigidity. The relation between the energy of a particle and its rigidity is given by

$$R = \frac{Pc}{Ze} \quad (1.11)$$

where P is the total momentum of the particle, Z is the atomic number of a particle, e is the charge of a proton, and R is the rigidity of a particle. In the energy region of interest (≥ 250 TeV) where energy of a particle is much larger than its rest mass, $Pc \approx E$.

There is a limit on the rigidity of a particle depending on the size of an accelerator. If the gyro-radius of a particle is larger than the size of an accelerator, then the particle escapes from the acceleration region. If a particle has this cutoff rigidity, then the maximum energy it can have depends on its charge. Helium can have twice the amount of energy than that of a proton, and iron can have 26 times the energy of the proton. This sequence of cutoff in energy based on the cutoff rigidity and a nuclear charge of a particle is called Peters cycle [3]. These cutoffs result in distinct features in the all-particle cosmic ray energy spectrum.

As mentioned before, the high energy cosmic ray spectrum has two distinct features, knee and ankle, dividing the spectrum into three distinct regions. The knee is around 3×10^{15} eV and the ankle is around 1×10^{18} eV. The knee is believed to be the beginning of the end of galactic cosmic rays [3] and the ankle is believed to be the end of galactic cosmic rays and the beginning of extragalactic cosmic rays [4]. The high energy cosmic ray spectrum can be divided into populations, usually three, and is modeled by

$$\phi_i(E) = \sum_{j=1}^n a_{ij} E^{-\gamma_{ij}} e^{-\frac{E}{Z_i R_{ij}}} \quad (1.12)$$

where j loops over n populations and i loops over five elemental groups, i.e., H, He, O, Si, and Fe. Fig 1.6 shows the modified H4a composition model [62] used for low energy spectrum analysis. This model has three populations. The third population consists of elements with cutoff rigidity 2 EV and 60 EV.

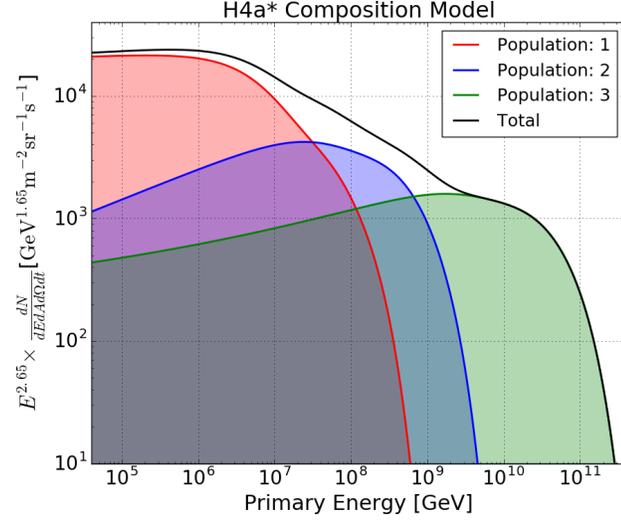


Figure 1.6: Three populations of the modified H4a composition model [62] and their sum. The third population consists of mixed elements with cutoff rigidity 2 EV and only protons with cutoff rigidity 60 EV.

1.5 Extensive Air Showers

Cosmic rays reaching the Earth’s atmosphere interact with air. At each interaction, the energy of the primary particle is shared among the newly formed secondary particles. There are muonic, electromagnetic, and hadronic components to an air shower. A hadron colliding with air produces other hadrons, pions, and kaons. A neutral pion decays into two photons. A photon produces an electron (e^-) and positron (e^+) by pair production. An electron then generates a photon by bremsstrahlung. This cycle repeats until the energy of e^\pm drops below the critical energy or is absorbed in the ground. A charged pion decays into a muon and a neutrino. A muon can penetrate through matter more than an electron or a photon. The electromagnetic case from a neutral pion decay is given by

$$\begin{aligned}\pi^0 &\rightarrow 2\gamma \\ \gamma &\rightarrow e^+ + e^- \\ e^\pm &\rightarrow \gamma + e^\pm\end{aligned}$$

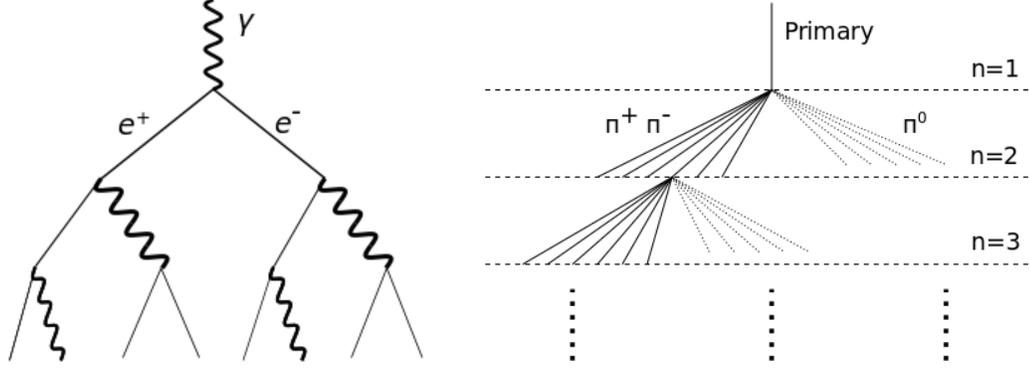
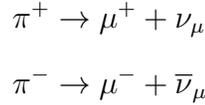


Figure 1.7: Left: Schematic of Heitler’s model for electromagnetic air shower. Right: Schematic of Heitler’s model for hadronic air shower.

And the hadronic case is given by the combination of the electromagnetic case and the following



Collision and decay of particles creates an extensive air shower of secondary particles. An electromagnetic shower consists of photons, electrons, and positrons. Photons produce electrons and positrons by pair-production. Electrons and positrons radiate photons of lower energy by Bremsstrahlung. A model that captures these processes is Heitler’s model [63]. A schematic of Heitler’s model is shown in Fig 1.7. For an air shower with primary energy E_0 , assuming Heitler’s model, shower maximum (X_{\max}) and the number of particles at X_{\max} for both electromagnetic and hadronic shower can be calculated. For an electromagnetic shower

$$\begin{aligned}X_{\max} &= \lambda_r \ln(E_0/E_c) \\ N(X_{\max}) &= E_0/E_c\end{aligned}$$

where λ_r is the radiation length in that medium and $E_c \approx 87 \text{ MeV}$ is the critical energy.

A study to expand Heitler’s model for hadronic showers was performed by J. Matthews [64]. In this modified model, the number of charged pions (π^\pm) produced

is N_{ch} and the number of neutral pions produced is $1/2N_{\text{ch}}$ in each interaction. The primary energy is also assumed to be equally distributed among π^0 , π^+ , and π^- in each interaction. The energy and shower maximum after n iterations becomes

$$E_{\pi^\pm} = \frac{E_0}{\left(\frac{3}{2}N_{\text{ch}}\right)^n}$$

$$X_{\text{max}}^{\text{p}} = X_0 + \lambda_r \ln\left[\frac{E_0}{3N_{\text{ch}}E_c}\right]$$

$$X_{\text{max}}^{\text{A}} = X_{\text{max}}^{\text{p}} - \lambda_r \ln A$$

where X_0 is the depth of first interaction, $X_{\text{max}}^{\text{p}}$ is the depth of shower maximum for proton shower, $X_{\text{max}}^{\text{A}}$ is the depth of shower maximum for element with atomic number A , and $\lambda_r \approx 37 \text{ g/cm}^2$ is the radiation length in air.

1.6 Cascade Equation

An approach to formulate a shower's profile is through the cascade equation. Cosmic rays collide with the atmosphere. Collision distributes the energy of the primary particle into many secondary particles. Depending on the energy and type, each of these secondary particles further collides or decays and produces more particles. This process continues until the energy of the secondary particle is completely lost or the particle hits the ground. A cascade of particles is formed through the process of collision and decay. The development of this cascade in the atmosphere is explained in detail in [1] and is described by

$$\frac{dN_i(E_i, X)}{dX} = -\frac{N_i(E_i, X)}{\lambda_i} - \frac{N_i(E_i, X)}{d_i} + \sum_{j=i}^J \int_E^\infty \frac{F_{ji}(E_i, E_j)}{E_i} \frac{N_j(E_j, X)}{\lambda_j} dE_j \quad (1.13)$$

where $N_i(E_i, X)dE_i$ is the flux of particle of type i at slant depth X with energy in between E and $E + dE$. λ_i is the interaction length of particle i in air. d_i is the decay length. $F_{ji}(E_i, E_j) = E_i \frac{dN(E_j, E_i)}{dE_i}$ is the particle yield function when particle of energy E_j collides with air and produce an outgoing particle of energy E_i .

The majority of cosmic rays are protons. The simplest form of cascade equation to solve is for nucleons (proton plus neutron). The cascade equation for a nucleon becomes

$$\frac{dN(E, X)}{dX} = -\frac{N(E, X)}{\lambda_N(E)} + \int_E^\infty \frac{F_{NN}(E, E')}{E} \frac{N(E', X)}{\lambda_N(E')} dE' \quad (1.14)$$

as the decay term is ignored. Nucleons interact before they decay because of a long decay time. If the radiation length (λ_N) is energy independent and inclusive cross sections ($F_{NN}(E, E')$) for pair production and Bremsstrahlung scale, the solution of the cascade equation subject to a power-law boundary condition is given by

$$N(E, X) = g(0)e^{-X/\Lambda} E^{-\gamma} \quad (1.15)$$

where Λ is the attenuation length and γ is the differential spectral index. $g(0)$ is the flux of nucleons at the top of the atmosphere. The attenuation length is given by

$$\frac{1}{\Lambda} = \frac{1 - Z_{NN}}{\lambda_N} \quad (1.16)$$

where Z is the spectrum-weighted inclusive cross section and is given by

$$Z = \int_0^1 x^{\gamma-1} F(x) dx \quad (1.17)$$

where x is the ratio of energy of secondary particles to primary particles.

The solution of Eq. 1.13, with the delta function at the energy of the nucleus boundary condition at the top of the atmosphere, is relevant for low energy spectrum analysis. The analytic solution with this boundary condition is discussed in detail in [1, 65]. For the low energy spectrum analysis, the Monte Carlo simulations provide solutions of Eq. 1.13 subject to the delta function boundary condition.

In a similar manner, the cascade equation can be expanded for photons, muons, neutrinos and other particles and detailed calculations are shown in [1].

1.7 Detection Method

Cosmic rays can be measured directly with a detector in space or on a balloon, or indirectly with detectors on the ground. Direct measurements measure composition

and energy of primary particles up to 100 TeV. As energy increases, the rate of events decreases. Therefore, a ground-based detector with a large effective area is required to collect high energy events in a reasonable time. The ground-based detector detects air showers created by high energy particles of unknown composition and energy. Therefore, a careful analysis is required to reconstruct these shower properties.

There are many experiments to detect cosmic rays. The Payload for Antimatter Exploration and Light-nuclei Astrophysics (PAMELA) [66] is a well-known satellite-based detector. As the name suggests, it detected antimatter and light nuclei particles from 2006 to 2016. Another satellite-borne detector, located on the International Space Station (ISS), is the Alpha Magnetic Spectrometer (AMS) [67]. Its primary goal is to detect antimatter and to search for dark matter. Well-known balloon-borne detectors are Cosmic Ray Energetics And Mass (CREAM) [68] and Advanced Thin Ionization Calorimeter (ATIC) [69].

There are many ground-based cosmic ray detectors. To name a few are IceTop [70], ARGO-YBJ [71], HAWC [72, 73], Cascade [74] (decommissioned in 2013), Tibet [75], Tunka [76], Telescope Array [77], and Auger [78]. ARGO-YBJ is a ground-based detector located in Tibet at an altitude of 4300 m that detects cosmic rays in the energy range of a few TeV to a few PeV [79]. HAWC is a ground-based gamma ray and cosmic ray detector located in Mexico at an altitude of 4100 m. The range of all particle cosmic ray energy spectrum published by the HAWC collaboration is from 10 TeV to 500 TeV [80]. Cascade was a ground-based detector in Germany that measured the cosmic-ray energy spectrum from 100 TeV to 100 PeV [81]. Tibet is a ground-based detector located in Tibet. It has measured all-particle cosmic rays energy spectrum from 112 TeV to 178 PeV [82, 83]. Tunka is a ground based detector located in Siberia. It can detect air showers from cosmic rays and gamma rays with energy in the range of hundreds of TeV to hundreds of PeV. Telescope array and Auger are ground based detectors that detect ultra high energy cosmic rays with energy higher than 100 PeV [84, 85]. Along with cosmic ray detectors, there are many ground-based gamma ray detectors. The major imaging atmospheric Cherenkov telescopes to detect

gamma rays are HESS in Namibia, MAGIC [21] on the Canary island of La Palma, and ‘Very Energetic Radiation Imaging Telescope Array System’ (VERITAS) [22] in Arizona. ‘Cherenkov Telescope Array’ (CTA) [86] is under construction and will be built in Spain and in Chile.

Data used in analyses for this dissertation is from IceTop [87], which is the surface component of the IceCube Neutrino Observatory [88]. A reason to build IceCube is to detect neutrinos. There are also other neutrino detectors around the world, like Super-Kamiokande [89], KM3NeT [90] etc. Neutrinos interact weakly and require a large detector to detect statistically significant number of them. This requirement was realized from the beginning of its detection in 1956 [91]. The initial attempt to build a large detector was in water and was called ‘Deep Underwater Muon and Neutrino Detector’ (DUMAND) [92]. On the South Pole, ‘Antartic Muon and Neutrino Detector Array’ (AMANDA) [93] was built in a clear ice which took data from 2000 to 2009. AMANDA was superseded by IceCube. For cosmic ray physics, there were two generations of scintillator based experiment called ‘South Pole Air Shower Experiment’ (SPASE) [94, 95] at the South Pole. IceTop, which uses the same *data acquisition* (DAQ) system as the rest of the IceCube detectors, replaced SPASE. The description of IceCube, focusing on IceTop, is discussed in detail in Chapter 2.

The major part of this dissertation is to lower the energy threshold of the IceTop cosmic ray detector. The motivation to lower the energy threshold of IceTop is to reduce the gap in energy between direct and indirect measurements and have an independent study to get energy spectrum around the knee region. The low energy all-particle cosmic ray spectrum contributes in the study of the cosmic ray composition that can be used to develop a more accurate cosmic ray composition model. The cosmic ray energy spectrum along with its composition can help us understand more about the physics behind its acceleration and propagation.

Chapter 2

DETECTOR

The IceCube Neutrino Observatory is a detector designed primarily to look for astrophysical neutrinos [96]. It is located at the geographic South Pole where there is a large volume of clear ice. Its construction was completed on December 18, 2010. The IceCube detector consists of two parts, InIce [88] and IceTop [87]. This chapter deals with both parts describing the mechanism used in IceCube to collect, record, and calibrate data with a focus on IceTop.

2.1 InIce

InIce extends from 1450 m to 2450 m below the surface occupying a volume of 1 km^3 of ice as shown in Fig 2.1. IceCube uses light detecting instruments called *Digital Optical Module* (DOM) [88] to detect light in clear ice. DOMs are discussed in more detail in Section 2.3. There are 86 vertical strings of wire each with 60 DOMs. The standard InIce consists of 78 strings deployed in a triangular grid with 125 m distance between the nearest strings. The distance between the consecutive DOMs in each of these strings is 17 m. InIce also has a denser region called DeepCore consisting of 8 strings with an average inter-string distance of 72 m. The first 10 DOMs of each DeepCore strings are placed at an interval of 10 m ending at 2000 m below the surface. The remaining 50 DOMs are placed at intervals of 7 m between 2100 m and 2450 m below the surface. The region between 2000 m to 2100 m below the surface has a layer of dust where DeepCore DOMs are not deployed.

IceCube was designed to detect astrophysical neutrinos with an energy in the range of TeV to few PeV. The DeepCore has denser geometry and has DOMs with higher efficiency. This combination helps lower the energy threshold to 10 GeV and

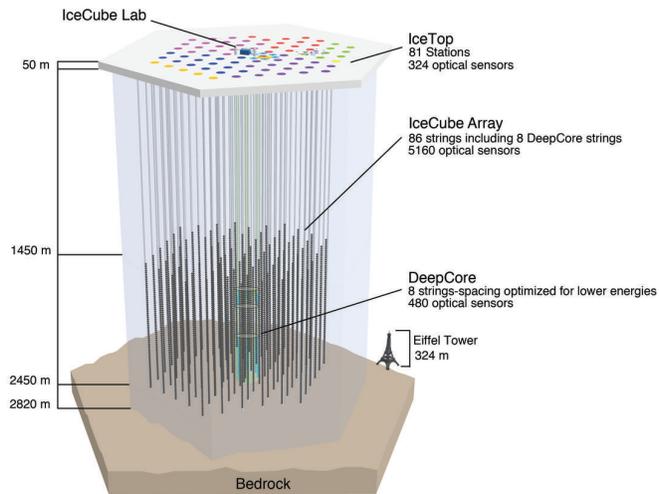


Figure 2.1: Schematic of the IceCube Neutrino Observatory showing IceTop and InIce. IceTop is at the surface of almost 3 km thick ice-sheet. InIce extends from 1450 m to 2450 m below the surface [88].

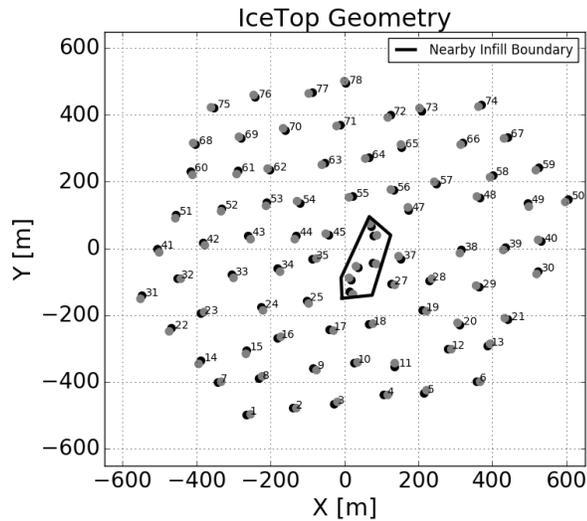


Figure 2.2: IceTop geometry showing exact position of all 162 tanks. Two tanks are placed in distance of 10 m and forms a station. Distance between nearby stations is around 125 m. Area surrounded by an irregular polygon at the center has nearby infill stations.

increases the probability of IceCube to detect neutrino oscillations, neutrinos from WIMP dark matter annihilation, and Galactic supernovae [97].

The mechanism involved in detecting neutrinos is as follows. Neutrinos interact via the weak force mediated by W^\pm for charged-current interaction and by Z^0 for neutral-current interaction. The charged-current reaction is $\nu_x + N \rightarrow x + X$ and the neutral-current reaction is $\nu_x + N \rightarrow \nu_x + X$. Neutrino interactions in IceCube have cascade and track-like topologies. The cascades are created by secondary leptons in ν_e and ν_τ charged-current interactions, by neutral-current interaction of all flavors of neutrino, and by hadronic fragmentation. Tracks are created by secondary muons in ν_μ charged-current interactions. Secondary charged particles produce photons if they are traveling faster than the speed of light in the ice or rapidly losing energy by Bremsstrahlung. DOMs detect these photons. Based on the position of hit DOMs, the time of the hit, and the amount of charge deposited, the direction and energy of neutrinos are reconstructed. The directional resolution of track-like events is less than a degree and that of cascade events is 10° to 15° [96].

2.2 IceTop

IceTop is the surface component of the IceCube Neutrino Observatory. It detects extensive air showers produced by high energy cosmic rays. Data collected by IceTop are primarily used to study the cosmic ray energy spectrum [98, 99, 100, 101] and the mass composition of primary particles [102], and to calibrate InIce [103]. It has also been used to look for PeV gamma rays [104], solar ground level enhancements (GLEs) [105], and transient events, like solar flares [105] or gamma ray bursts. There is an ongoing effort to use IceTop as a veto to get rid of backgrounds in astrophysical neutrino search.

IceTop is located at the South Pole at an altitude of 2835 m above sea level. IceTop consists of 162 tanks distributed in 81 stations spread over an area of 1 km^2 . The position of all tanks is shown in Fig 2.2. Each station has two tanks separated by 10 m. Stations are arranged in a triangular grid and two nearby stations are around

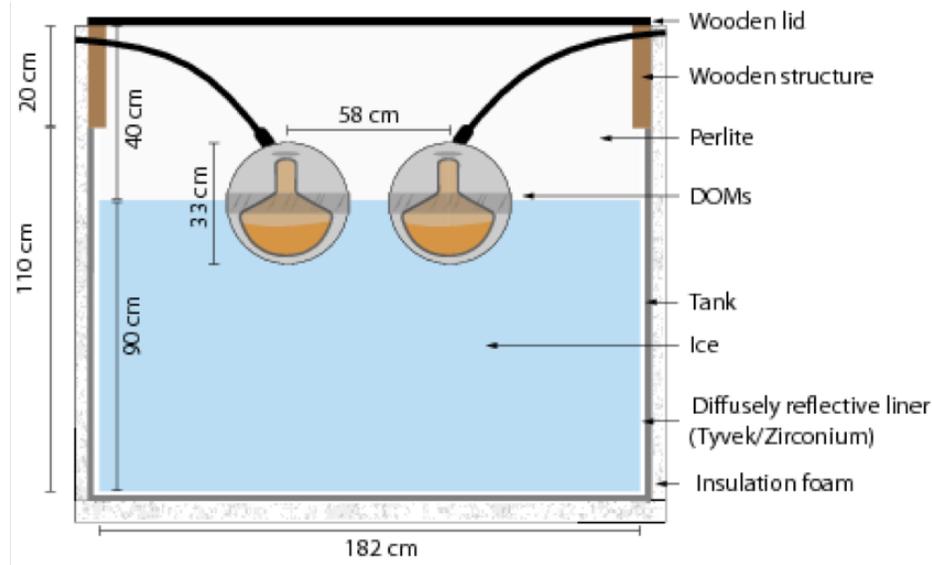


Figure 2.3: Schematic of an IceTop tank.

125 m apart. IceTop also has a dense infill array where the distance between nearby stations is approximately 60 m.

All tanks are made of 6 mm thick polyethylene by a rotational molding technique. A tank is a cylindrical shaped container with an inner diameter of 1.82 m and a height of 1.1 m. A schematic of an IceTop tank is shown in Fig 2.3. In most of the tanks, the inner layer consists of an additional 4 mm of zirconium-dioxide as shown in Fig 2.4. Only 12 tanks have a bag of Tyvek loosely covering the tank wall. All tanks were filled with water and frozen using a Freeze Control Unit to form the clearest ice possible. Each tank is equipped with two DOMs with different gains to detect light emitted by charged particles in clear ice. DOMs were partially immersed in water with the photo-multiplier tubes (PMTs) facing downward. Both DOMs of a tank were supported by a wooden structure with a joist and a lid until the water froze. Once the ice was formed, the space between the ice surface and the wooden lid was filled with perlite and then the lid was closed permanently.

The standard IceTop geometry is designed to detect cosmic rays in the energy range from PeV to EeV. The dense infill array is used to detect cosmic rays with



Figure 2.4: A piece of IceTop tank and a US nickel for scale. The black top is the outer polyethylene layer of a tank. The white bottom layer is the layer of reflective zirconium.

comparatively lower energy. The reason for having two tanks in a station is to suppress background low energy showers at the rate of ~ 2 kHz hitting only one tank. If both tanks of a station are hit within $1 \mu\text{s}$, then the “Hard Local Coincidence” (HLC) condition is fulfilled and full waveform readout is initiated. If only one tank of a station is hit then a “Soft Local Coincidence” (SLC) is said to be fulfilled, where the waveform is ignored and only integrated charge and timestamp is recorded to save disk space. Two tanks in a station are also used to cross-check reconstruction parameters by splitting the detector into two arrays, each containing only one tank from a station. The reason for having two DOMs in a tank is to determine tank’s signal fluctuations by setting both of them to the same gain during special calibration runs, and to use one if the other one is dead. In order for DOMs to have enough light to trigger on, the wall of the tanks should be reflective. Both zirconium dioxide and Tyvek are good reflectors of visible light. The reason to fill tanks with perlite is for thermal insulation and to exclude light from outside.

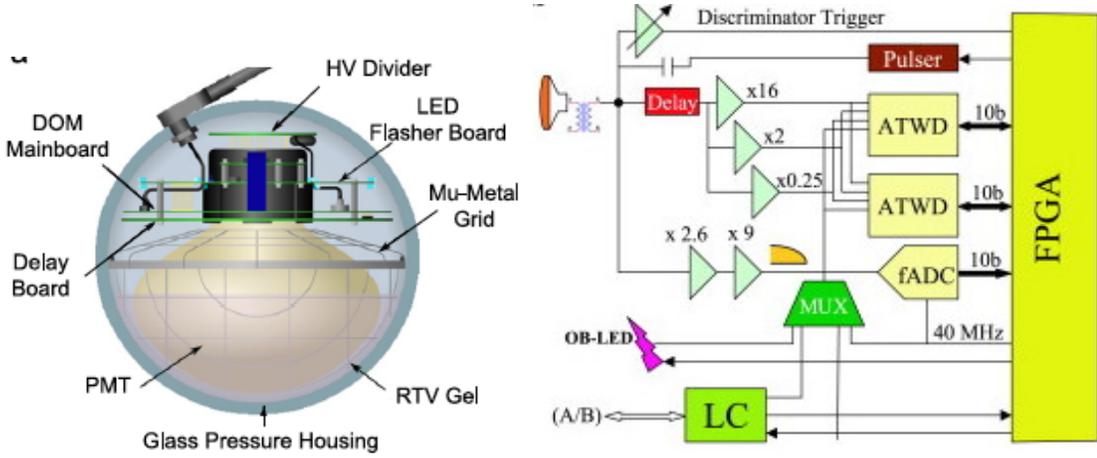


Figure 2.5: Left: Schematic of a DOM. Right: A block diagram of the DOM mainboard. Both figures are from [87]

2.3 Digital Optical Module (DOM)

Digital Optical Module (PMT and other electronics) (DOM) is the fundamental detection unit for InIce and IceTop. A DOM is a glass pressure sphere of 33 cm diameter containing tools to detect, analyze, digitize, and communicate. A *photo-multiplier tube* (PMT) of a DOM is a photon detector that multiplies the signal produced by incident photons. The *Field Programmable Gate Array* (FPGA) controls data taking, triggering, digitizing and communicating with the data collection and processing center called the *IceCube Lab* (ICL). An Analog Transient Waveform Digitizer (ATWD) is the electronics where a signal is digitized. A DOM also has a free-running 20 MHz oscillator used as a local timer. This oscillator is regularly calibrated with the master clock in the ICL. Schematics of a DOM and its electronics is shown in Fig 2.5.

Charged particles entering IceTop tanks produce Cherenkov light if the speed of the particle is greater than the speed of light in ice. A PMT captures these photons. Due to the photo-electric effect, electrons from the valence shell of a metal used in the PMT entrance window are knocked out. The number of electrons multiplies inside the PMT depending on its gain. Refer to [106] for a detailed working mechanism of a PMT. At the end of a PMT is a transformer to decouple signal from PMT's high voltage.

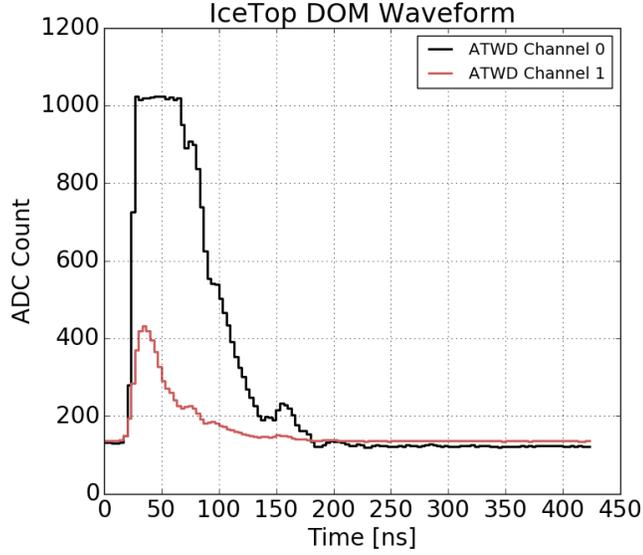


Figure 2.6: An example waveform. Channel 0 of ATWD is saturated so channel 1 of lower gain is used for signal digitization. The digitized value of waveform of each ATWD bin is called an ADC count.

The decoupled signal is split for checking signal’s strength and for recording. If the signal voltage passes a certain programmable threshold and if the local coincidence (LC) condition is fulfilled, then the ATWD starts sampling the signal in a bin width of 3.33 ns. The ATWD records 128 total samples and takes 427 ns to sample a signal. Before recording starts, there is an artificial delay of 75 ns for the FPGA to make decisions. There are two ATWDs to minimize dead-time. Each ATWD has three channels with different gains of 16, 2, and 0.25 for recording signals. In each channel at the same time, the signal is sampled every 3.33 ns 128 times and information is stored to be written on a disk. The digitization process begins with a 10-bit precision. The bin-wise record of a signal is called a waveform. Waveform information is digitized one after another in each channel. The lower gain channels are only digitized if in the higher gain channel any bin is above 768 counts (mid-value of 2^9 and 2^{10}). It takes 30 μs to digitize information from an individual channel and a total of 100 μs if it has to digitize all three channels.

Fig 2.6 is an example waveform recorded by an IceTop DOM after digitizing

it on bin-by-bin basis. The digitized value of waveform of each ATWD bin is called an ADC count. The highest gain channel of the ATWD is saturated, so the lower gain channel records the complete waveform. A waveform is used to extract the total pulse charge and the time of the hit. The unsaturated waveform from a lower gain ATWD will be used among the two waveforms shown in Fig 2.6. An ADC count for each ATWD bin is converted to voltage. The conversion relation is found by linearly fitting ADC amplitudes after varying the ATWD reference voltage for each ATWD bin during calibration [87]. Also, the transit time for a signal to travel through a PMT and the DOM electronics is subtracted. Then a droop effect due to the transformer is corrected¹. Now the total charge of a pulse is the sum of a calibrated waveform in units of photo-electrons (pe). The pulse time is the time where line passing between 10% and 90% of the leading edge of the waveform intersects the baseline. Usually, charge and time from high-gain DOM are used unless it is saturated.

2.4 Data Transfer

A digitized waveform is temporarily stored in the local DOM memory and then transferred to a DOMHub using a single twisted pair of wires. Information stored in a DOM is sent upon request by the DOMHub in regular intervals. One DOMHub with one DOM readout card (DOR) is responsible for eight stations with 32 DOMs. Once received, the timestamps of each hit are calibrated and then sorted by calibrated time of the hit. The time calibration is discussed in detail in section 2.5.3. At this point, many trigger conditions are checked to form and collect events that can be used to do physics. Once all trigger conditions are checked, a global trigger merges all overlapping hits to form an event. The result is stored in a DAQ data format and sent for online processing and filtering (PnF). The total data collected daily by IceCube is around 1 TeraByte.

¹ Droop is the part of waveform that lies below the baseline. Baseline is the average of intercepts in ADC count to voltage relation from all ATWD bins (see Section 2.5.1). The detailed explanation on how to correct droop of a waveform is discussed in [87].

Data are collected at the South Pole and are transferred to Madison, WI for analysis. Data transfer is done via satellite. The use of satellite has a limitation on bandwidth. Data is reduced for satellite transmission using filter conditions designed for specific physics goals. IceTop has IceTopSTA3, IceTopSTA5, IceTop_InFill_STA3, and InIceSMT_IceTopCoin filters that will be discussed in more detail in Chapter 3. For low energy spectrum analysis, a new filter called IceTop_InFill_STA2, based on a new IceTop_Volume trigger, is implemented. Waveform information is saved and is transferred only for events passing IceTopSTA5 filter. The total charge and the leading edge time of hits are transferred for events passing other filters. The data, in SuperDST format, is transmitted via satellite. SuperDST is an extremely compressed data format used to save satellite bandwidth while transferring data from the south pole to computers in Madison, Wisconsin.

2.5 Calibration

A signal is calibrated to obtain charge in units of *photo-electron* (pe) and the time relative to the master clock at the ICL for all DOMs throughout the detector. A charge in pe is further converted into a tank independent charge unit called *Vertical Equivalent Muon* (VEM). One VEM is the signal in pe a DOM detects due to an energetic muon traveling vertically through a tank. An energetic muon deposits equal amount of energy per unit length while traveling through a matter [107]. A software called DOMCal is used to calibrate charge and time of each DOM, and a software called VEMCal is used to calibrate signal charge in pe to VEM for each DOM. A method (discussed in Section 2.5.3) called ‘Reciprocal Active Pulsing Calibration’ (RAPCal) is used to calibrate time.

2.5.1 Calibration of DOM electronics and PMT

The calibration of the front-end amplifier, the voltage measurement for each ATWD bin, sampling speed of ATWD, and PMT gain is performed during a special calibration run with a software tool called DOMCal. IceTop is exposed to temperature

fluctuations, so the DOMCal procedure is performed once a month. DOMCal for InIce is run once a year. DOMCal is run in a half detector configuration while the remaining half still takes physics data. The reference quantities for DOM calibration are a known charge from an electronic pulser circuit, a DC bias voltage, a 20 MHz oscillator for timing and a signal generated by a single photo-electron produced by electronics noise or a low-luminosity LED.

An ATWD signal from each bin is digitized to ADC count using 128 Wilkinson ADCs [87]. The charge calibration converts the ADC counts of each ATWD bin to voltages. The process begins by generating signals with various DC bias voltage. ADC counts are recorded for each ATWD bin for each DC bias voltage. The result of ADC count vs. voltage for each bin is fitted with a linear function, resulting in a slope and an intercept for each ATWD bin. These parameters are used to convert ADC counts to voltage for each ATWD bin. The intercept of the fit gives the value of ADC counts when the input voltage is zero. This value is called a pedestal. The average of the pedestal value for all bins is called baseline and it is subtracted from the waveform to calculate the total charge of a pulse.

The ATWD channel with gain 16 is calibrated first, using single photo-electron like pulses generated by the pulser. The charge from a pulser is known and is compared with the measured pulse charge on each ATWD. The remaining two low gain channels are calibrated based on this high gain channel using LED light injected into the PMT. A single photo-electron pulse from cathode noise is passed through the DOM to calibrate PMT gain. The charge distribution obtained from these pulses is described by a Gaussian single photo-electron peak and an exponential contribution at low charges. The charge corresponding to the peak is mapped to the charge from a single photo-electron to calculate the PMT gain.

2.5.2 VEMCal

VEMCal is a calibration procedure which calculates the conversion factor from charge deposited in each DOM in pe to VEM. It is performed once a week for all HG

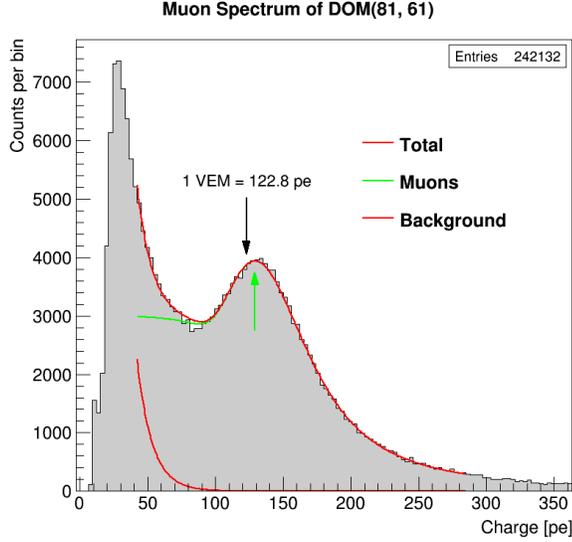


Figure 2.7: Experimental muon spectrum on DOM 81-61 fitted with Eq 2.1.

DOMs of IceTop. SLC hits following every 8192nd (2^{13}) HLC hit are used for VEM calibration. The rate of these ‘minimum bias’ hits is around 0.2 Hz per DOM. Data from one week is required to obtain reasonable statistics for the calibration. These hits are extracted during the regular data processing at the South Pole and are transferred to the North with the normal physics data stream. Using this data, a charge histogram is produced and stored daily for each DOM. These histograms are added once a week and fitted with an exponential function for the electromagnetic part and with a combination of a normalized Landau and a Fermi function for the muonic part. The function used for fitting is given by

$$f(x) = p_0 \left[\frac{1.85}{p_1 e^{\left(\frac{x-p_1}{p_2} + 1\right)}} + \text{Landau}(x, p_1, p_2) \right] + p_3 e^{(p_4 x)} \quad (2.1)$$

In this expression x is the charge in pe. p_0 , p_1 , p_2 , p_3 , and p_4 are parameters of the fit. The pe value where muon contribution is maximum, i.e. 95% of the muon peak as shown in Fig 2.7, is defined as 1 VEM.

Data are collected only from high-gain (HG) DOMs for VEMCal. So the relation between the charge deposited on a low-gain (LG) DOM to a VEM value has to be

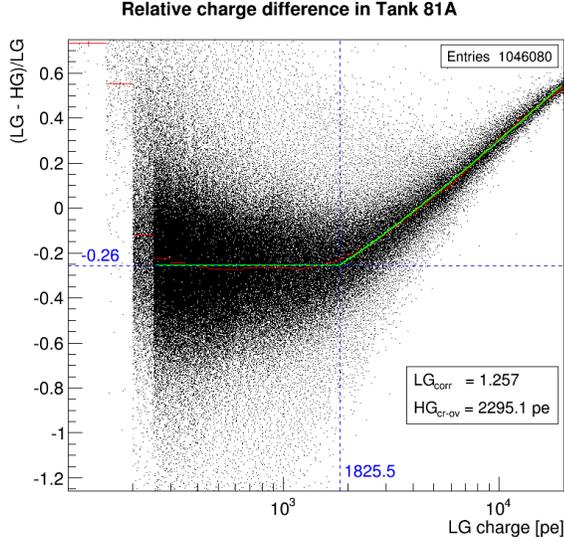


Figure 2.8: Relative charge difference of low-gain DOM and high-gain DOM of tank 81A.

derived based on the corresponding HG DOM. The charge deposited in both HG and LG DOMs of a tank should be the same, on average. So a charge on the LG DOM is adjusted to make it the same as on the HG DOM in the overlap region. As an example, the LG DOM seems to have a lower pe value for tank 81A in the overlap region, as shown in Fig 2.8. The lower value can be due to the uneven distribution of light in a tank or the difference in quantum efficiency between the PMTs. Based on Fig 2.7, 1 VEM for HG DOM of 81A corresponds to 122.8 pe. For the same tank, the correction factor for LG DOM is 1.257. Hence, 1 VEM for LG DOM of 81A is $122.8/1.257=97.69$ pe. This method is repeated for all LG DOMs. Refer to [87] for a detailed description.

2.5.3 RAPCal

The *Reciprocal Active Pulsing Calibration* (RAPCal) [88] is a process to synchronize all clocks used in IceCube. It converts DOM hit time to the ICL time and eventually to the Coordinated Universal Time (UTC). There is a clock at each DOM and the master clock at the ICL. The time at *DOM readout card* (DOR) is based on the time at the master clock. A pulse is exchanged between a DOM and a DOR.

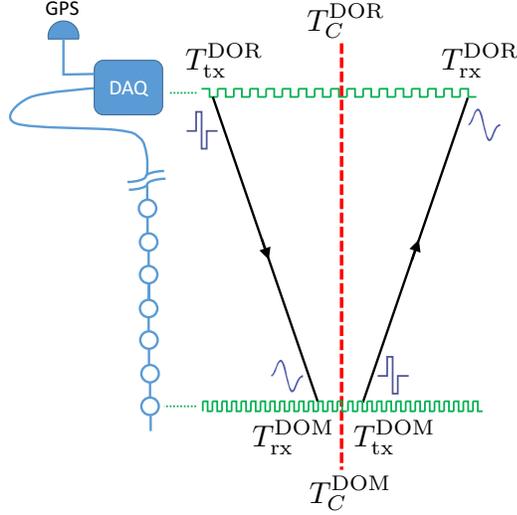


Figure 2.9: Diagram of the RAPCal time calibration scheme. Figure is from [88].

This exchange happens about once every second when there is no data transmission in power/communication wire pair. Let us assume, DOR transmits a pulse at time T_{tx}^{DOR} and a DOM receives it at time T_{rx}^{DOM} . After receiving a signal, a DOM transmits a similar pulse back to the same DOR. The DOM sends the pulse at time T_{tx}^{DOM} and received by the DOR at time T_{rx}^{DOR} . Time used for DOR and DOM is their respective local time.

The length of wire these two pulses travel along is the same. The midpoint time T_C for both DOM and DOR is given by²

$$T_C = \frac{T_{tx} + T_{rx}}{2}$$

From the symmetry of the problem as shown in Fig 2.9, T_C^{DOM} and T_C^{DOR} is the single point in time for both clocks and should be equal if the clocks were synchronized.

² A received pulse can have a droop and the time associated with it needs to be corrected before calculating the mid time. The detailed description of how time associated with droop in received pulses is corrected is discussed in [88].

A RAPCal is done approximately every second for each DOM. T_C^{DOM} and T_C^{DOR} are not always equal as expected. This small discrepancy in clock time, if any, can be seen by calculating ϵ using two consecutive RAPCal results, i.e

$$1 + \epsilon = \frac{T_{C,2}^{\text{DOR}} - T_{C,1}^{\text{DOR}}}{T_{C,2}^{\text{DOM}} - T_{C,1}^{\text{DOM}}}$$

$\epsilon = 0$ if there is no discrepancy. There is some discrepancy and the median magnitude of ϵ is 1.34×10^{-6} [88]. Finally, to convert hit timestamp at a DOM to a UTC time, the following relation is used.

$$UTC(t) = (1 + \epsilon)(t - T_C^{\text{DOM}}) + T_C^{\text{DOR}} + \Delta$$

where Δ is the UTC time offset and is provided by the master clock.

Chapter 3

NEW TRIGGER AND FILTER

One of the goals of this dissertation is to lower the energy threshold of the IceTop energy spectrum to overlap with the spectra from direct measurements. Low energy events hit a fewer number of IceTop stations separated by small distance. In an effort to collect these low energy events by IceTop, a new trigger and filter is developed and implemented requiring hits on two or more stations.

The widely used IceTop trigger is the IceTop Simple Multiplicity Trigger (IceTopSMT). It selects events with at least 6 HLC hits within $5 \mu\text{s}$. Once this condition is fulfilled, the time window of $10 \mu\text{s}$ before the first hit and $10 \mu\text{s}$ after the last hit of the 6 HLC hits is opened to collect data. Based on the IceTopSMT trigger, the IceTopSTA3 filter collects 3-station events and IceTopSTA5 filter collects 5-station events¹. All IceTop filters based on the IceTopSMT trigger are described in detail in Table 3.1. To study IceTop and InIce coincident events, the InIceSMT_IceTopCoincidence filter is implemented. This filter is based on the InIceSMT trigger and requires one HLC hit in IceTop.

The IceTopSTA5 filter has been used for the all-particle cosmic ray energy spectrum from a few PeV to a few EeV [98, 99, 100, 101]. The low energy thresholds using these events is $\sim 2 \text{ PeV}$. An intermediate study was conducted using IceTop_InFill_STA3, which lowered the cosmic ray energy spectrum to 500 TeV with an assumption that the cosmic ray flux is purely proton and 1 PeV with an assumption that the cosmic ray flux is purely iron [108]. To lower the energy threshold to the direct

¹ 3-station event has hits on at least three IceTop stations and 5-station event has hits on 5 or more IceTop stations. Similarly, two-station event has hits on at least two nearby infill stations.

measurement realm, a new trigger and filter was needed to collect low energy events with two infill stations hit.

Table 3.1: All filter conditions before the addition of two-station trigger and filter. SMT6 for IceTopSMT, STA3 for IceTopSTA3_13, STA5 for IceTopSTA5_13, IIT for InIceSMT_IceTopCoincidence_13, and InFill3 for IceTop_InFill_STA3_13 is used. IceTopSMT6 trigger is passed on events for first three filters. InIceSMT8 trigger is passed on events for InIceSMT_IceTopCoincidence_13 filter.

Tag	Condition	sDst		Filter	
		pre-scale	Rate [Hz]	pre-scale	Rate[Hz]
STA3	$N_{\text{station}} \geq 3$	1	22.3	10	2.2
STA5	$N_{\text{station}} \geq 5$	1	6.1	1	6.1
InFill3	$N_{\text{infill_station}} \geq 3$	1	3.7	10	0.37
IIT	$N_{\text{IT_HLC}} \geq 1$	1	100	100	1

A study that was done to estimate the rate of events and the additional bandwidth required to transfer data from the South Pole to Madison due to the new trigger and filter proposal is discussed in section 3.1. The new IceTop volume trigger and filter is defined in section 3.2. The performance evaluation of the new trigger and filter is discussed in section 3.3.

3.1 Background Study

Low energy events with less than 5 stations hit are numerous and their data demand more satellite bandwidth. Full waveform information is retrieved only for events when 5 or more stations are hit. The low energy threshold of all-particle cosmic ray energy spectrum using at least 3 infill stations events is 1 PeV [108]. In order to lower the energy threshold to 100 TeV, events with fewer than three stations hit are required. Therefore, a new IceTop trigger is implemented to select events with hits on 2 or more stations within a close distance.

A study was performed using Fixed Rate Trigger (FRT) data to estimate the rate of events and the additional bandwidth required to bring them to the North. FRT

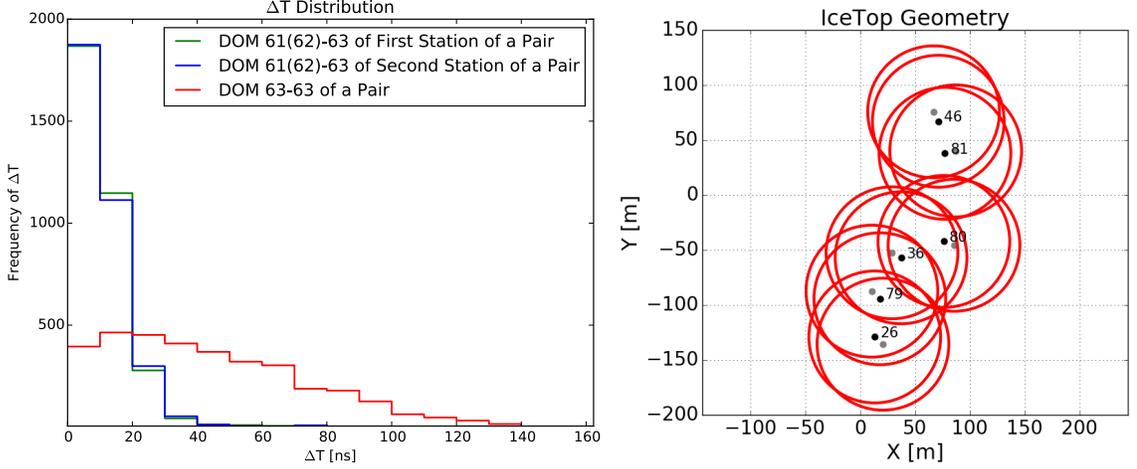


Figure 3.1: Left: The distribution of time difference between hits on DOMs 61 or 62 and 63 of each tank of a station is shown in blue and green. The distribution of time difference between hits on DOM 63 of each station of a nearby infill station pair is shown in red. Right: Six infill stations which form four nearby infill stations pairs are shown along with a circle around each tank of radius 60m.

data is a form of ‘minimum bias’ data, in which the entire detector is read out for 10 ms every 5 minutes. The FRT data from July to December of 2013 was used.

A 10 ms time window can result in many events in a single data frame. Therefore, event-splitting was performed to separate events for counting. All hits in a frame were sorted by the time of the hit. If the time difference between two consecutive sorted times was more than 420 ns, the event was split and the hit was pushed to the next possible event where it had to pass through the same selection process.

In order to develop a ‘two stations hit’ trigger, the maximum time difference required to generate a hit in one of the four pairs of nearby infill stations was studied. The left plot of Fig 3.1 shows the distribution of time difference between hits in tanks of a station and hits in tanks of a nearby station pairs. The time difference between all hits was less than 150 ns. Therefore, a 200 ns time window was used to define a trigger. The distance that light travels in 200 ns is 60 m, which is larger than the distance between the station pairs listed in Table 3.2.

As a trigger condition for an event, there must be hits on at least four high gain

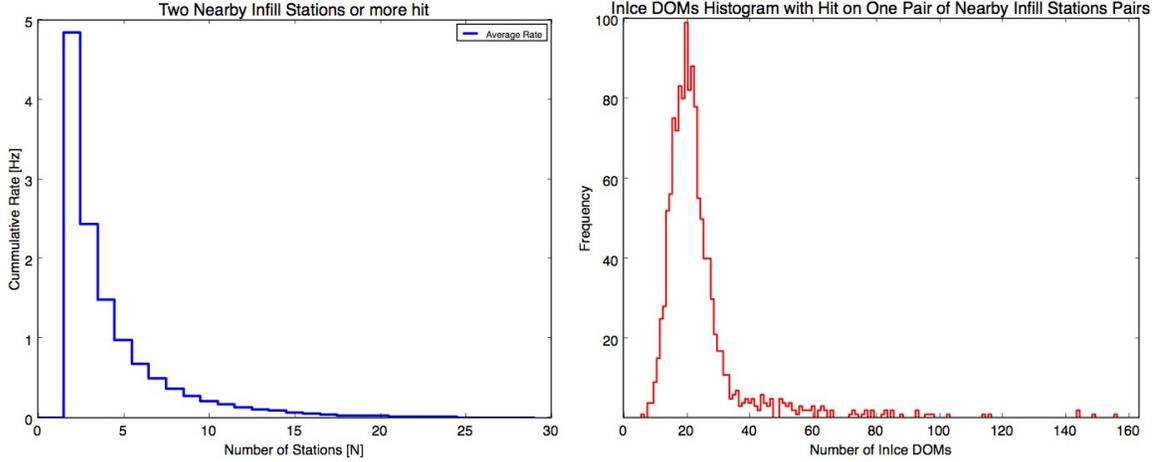


Figure 3.2: Left: Cumulative rate with hit on at least one nearby infill station pair. Data used is 2013 FRT from July to December. Right: Histogram of InIce DOMs that are hit along with four high gain DOMs hit in IceTop.

DOMs in any pair of nearby infill stations within 200 ns. The number of events that pass the trigger criteria in 10 ms time window and the number of stations hit in each event were estimated using FRT data. These events were used to estimate the new trigger rate.

The rate is defined as the number of events, divided by the duration of data-taking. The left plot of Fig 3.2 shows the cumulative rate of events as a function of number of stations. These stations must include at least two stations from nearby infill pairs shown in Table 3.2. The rate of two-station² events is 4.84 Hz and the rate for three stations hit events is 2.44 Hz. The three stations hit events can also be collected by IceTop_SMT6 trigger and IceTop_InFill_STA3 filter. Hence, the rate of events that hit only two infill stations is $4.84 - 2.44 \text{ Hz} = 2.4 \text{ Hz}$. The total online bandwidth used by the cosmic ray working group as listed in 2014-2015 TFT report is 5.27 GB/day. The additional rate of 2.4 events/s corresponds to 39.4 MB/day assuming, on average, 40 InIce DOMs hit for each two-station events. This addition was an 0.75 % increase

² two-station event has hits on at least one pair of IceTop infill stations listed in Table 3.2

in the total bandwidth required by the cosmic ray working group.

3.2 IceTop Volume Trigger

For the low energy spectrum analysis, a new trigger and filter has been implemented since May 20, 2016 to collect two-station events. The name of the trigger is “ICETOP_VOLUME” and the name of the filter is “IceTop_InFill_STA2_16”. The filter was renamed as “IceTop_InFill_STA2_17” since the 2017 data-taking season. Events that pass trigger and filter are stored in superDST data format and are sent to Madison via satellite.

The trigger algorithm is loosely described while estimating rate using FRT data. Each DOM in IceCube is autonomous and works independently. The new trigger is formed by counting the number of DOMs that are hit within a cylinder defined by a 60 m radius and a 10 m height in 200 ns time. This type of trigger is called a volume trigger. There are 2 HG DOMs in a station, one in each tank. Four nearby infill pairs have 6 stations which have 12 high gain DOMs in total (see Table 3.2). For an event to pass the trigger condition, it has to have HLC hits on at least 4 HG DOMs out of the 12 HG DOMs within 60 m radius and 200 ns time window. The 60 m radius circle around each tank of 6 stations is shown in the right plot of Fig 3.1. Once the trigger condition is fulfilled, the readout window starts $10\mu\text{s}$ before and after the first and last of the 4 HG DOMs hit. All triggered events automatically pass the filter condition and are sent to Madison, Wisconsin for analysis.

Table 3.2: Four pairs of nearby infill stations and distance between each pair in meter.

Stations	Distance [m]
46, 81	34.9
36, 80	48.9
36, 79	40.7
79, 26	41.6

3.3 Trigger/Filter Performance

The rate of events hitting only two stations was predicted to be 2.4 Hz using FRT data from 2013. The rate of events with only two stations hit using IceTop-Volume trigger on 2016 data is around 2 Hz, as expected. This shows that the new trigger/filter performed as expected. To test the performance of the trigger/filter every year, the rate of various measured quantities, like number of stations, total charge per event, and charge on hit tank, is monitored. The IceCube detector collects data in 8-hour blocks and tags each block of data with a run number. To test the performance of all triggers and filters, 3 test runs are performed in a 24 hour time period once a year. These test runs are compared with a previously taken reference run. Fig 3.3 shows the differential rate of number of stations hit, a histogram of total charge per event, and a histogram of charge per tank for two-station events.

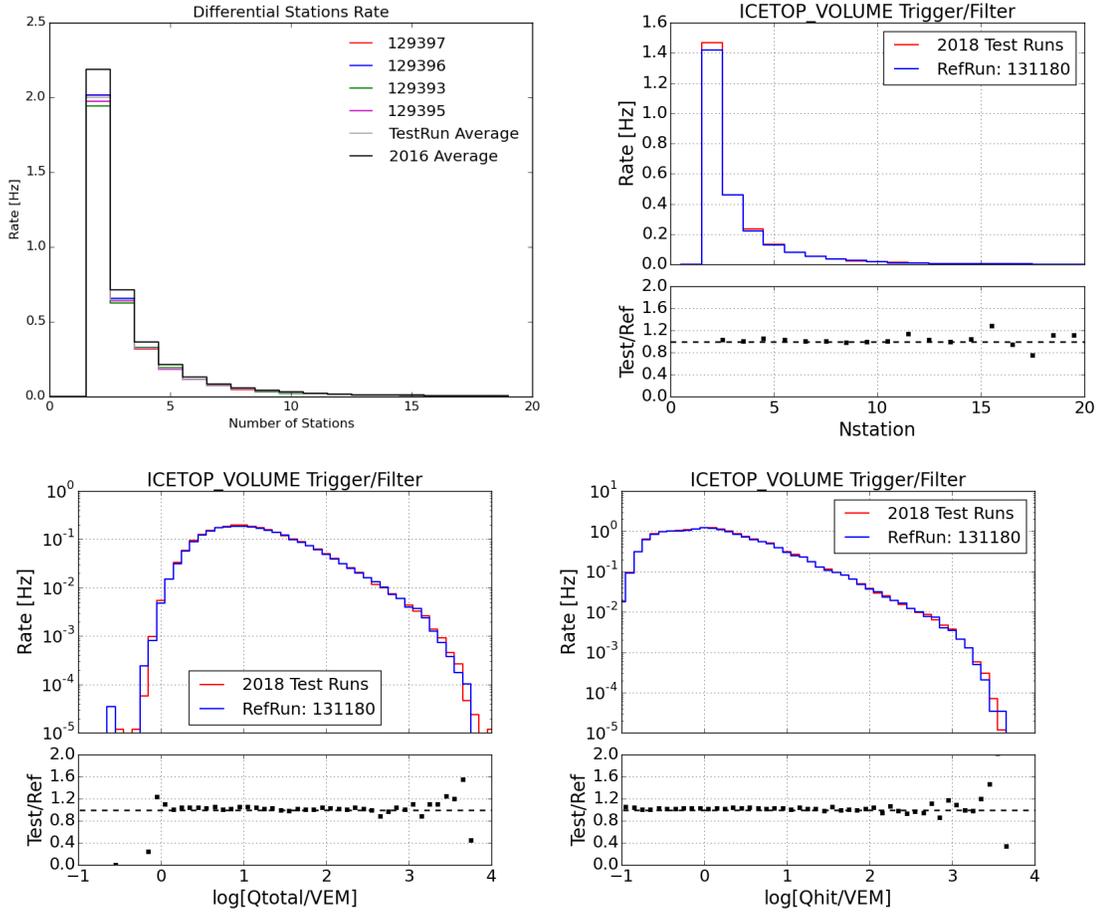


Figure 3.3: Left and right plots on top show differential station rate from 2016, 2017 and 2018 data. From 2016 data, the average rate of all runs, and from 2017 and 2018 data, rate for test runs is shown. The bottom left plot shows Q_{total} distribution for the 2017 reference run and the 2018 24hr test run. The bottom right plot shows Q_{hit} distribution for the 2017 reference run and the 2018 24hr test run. All four plots show that Trigger/Filter are working as expected.

Chapter 4

SIMULATION

One of the disadvantages of a ground-based detector is not knowing the energy of the cosmic ray causing an air shower. Knowledge of the primary has to be obtained from measured shower properties at the ground. To develop a method for shower reconstruction and to know its accuracy, simulations are used. Simulation depicts an air shower from a particle with a known type, position, direction, and energy. Each simulated shower produces many particles that reach the ground, which must then be propagated through a model of the IceTop tanks. To calculate the atmospheric cascade, a standard simulation package called CORSIKA [109] is used. For simulation of propagation of particles through the tanks, Geant4 [110] is used.

The IceCube collaboration has a team to produce standard CORSIKA simulations in order to optimize computational resources by avoiding repetition of work. The standard CORSIKA simulations of proton, helium, oxygen, and iron primaries ranging from 10 TeV to 25 PeV in energy are used for this analysis. To increase statistics, the same CORSIKA shower is used multiple times (re-sample) by changing its core position during detector simulation. After re-sampling, approximately one hundred thousand showers for each $0.1 \log_{10} E_{\text{primary}}$ bins within 0° to 65° zenith angle are produced. The zenith angle of all primary particles is from 0° to 65° , except for Helium and Oxygen with energy 10 TeV to 100 TeV where the zenith angle is from 0° to 40° . Table 4.1 is the list of all CORSIKA datasets of four primary particles with information on zenith range, energy range, hadronic interaction model, atmosphere, and the number of events in each energy bin. Sibyll2.1 [111] is used as the base hadronic interaction model for the low energy spectrum analysis. CORSIKA showers with QGSJetII-04 [112] as hadronic interaction model are also produced with 10%

of the statistics compared to that of Sibyll 2.1. The QGSJetII-04 model is used to do a parallel study to compare its result with that of Sibyll2.1. For the South Pole, atmospheric model for July 1, 1997 (see Table A.1) is used for showers with energy 10 TeV to 100 TeV and average April atmosphere (see Table 4.2) is used for showers with energy higher than 100 TeV.

Table 4.1: List of CORSIKA datasets that are used in this analysis. Hadronic interaction models used are Sibyll2.1 and QGSJetII-04. Two analyses are performed; one with Sibyll 2.1 and the other with QGSJetII-04 as hadronic interaction models.

DataSet	Element	$\log_{10}[E/\text{GeV}]$	Zenith	Had. Model	Atmos.	Nevt/ $0.1\log_{10}E$
10410	H	5-8	0°-65°	Sybill2.1	AprAvg	100,000
11663	He	5-8	0°-65°	Sybill2.1	AprAvg	100,000
12605	O	5-8	0°-65°	Sybill2.1	AprAvg	100,000
10889	Fe	5-8	0°-65°	Sybill2.1	AprAvg	100,000
9508	H	4-5	0°-65°	Sybill2.1	Jul 1, 1997	100,000
7362	He	4-5	0°-40°	Sybill2.1	Jul 1, 1997	70,000
7364	O	4-5	0°-40°	Sybill2.1	Jul 1, 1997	70,000
9614	Fe	4-5	0°-65°	Sybill2.1	Jul 1, 1997	100,000
10951	H	5-8	0°-65°	QGSJet II-04	AprAvg	10,000
12583	He	5-8	0°-65°	QGSJet II-04	AprAvg	10,000
12584	O	5-8	0°-65°	QGSJet II-04	AprAvg	10,000
10954	Fe	5-8	0°-65°	QGSJet II-04	AprAvg	10,000
20252	H	4-5	0°-65°	QGSJet II-04	AprAvg	10,000
20253	He	4-5	0°-65°	QGSJet II-04	AprAvg	10,000
20254	O	4-5	0°-65°	QGSJet II-04	AprAvg	10,000
20255	Fe	4-5	0°-65°	QGSJet II-04	AprAvg	10,000

In CORSIKA, the density variation of the atmosphere with altitude is modeled in 5 layers. The overburden of the atmosphere in these 5 layers is given by

$$\begin{aligned}
 T(h) &= a_i + b_i \exp\left(-\frac{h}{c_i}\right) & i = 1, 2, 3, 4 \\
 T(h) &= a_5 - b_5 \frac{h}{c_5}
 \end{aligned}
 \tag{4.1}$$

where h is the altitude above the sea level in units of cm, $T(h)$ is the overburden at altitude h in unit of g/cm^2 . The CORSIKA atmosphere (ATMOD11–14) shows unphysical behavior if compared with data collected by detectors on the AMRC balloon and AIRS satellite. The top two plots of Fig 4.1 shows some unphysical jumps in temperature and density profile at heights of 4, 10 and 40 km. There is also a significant difference in atmospheric depth profile between data and CORSIKA simulations, as shown in the bottom plot of Fig 4.1. This shows the difference between atmospheric depth in the July 01, 1997 atmosphere model provided by CORSIKA and the daily July 2010 atmospheric depth measured by AMRC. A study was conducted by Sam DeRidder to have a more realistic model for each month. The following parameters were proposed for the first line of Eq. 4.1 for the average April atmosphere.

Table 4.2: Parameters for April Atmosphere used in the simulation. Only first line of Eq. 4.1 is used for all five layers. h is in unit of km and IceTop is located in Layer1.

Parameter	Layer1	Layer2	Layer3	Layer4	Layer5
a	-69.7259	-2.79781	0.262692	-8.41695e-05	0.00207722
b	1111.7	1128.64	1413.98	587.688	-
c	766099	641716	588082	693300	5.4303203e9
h	0 - 7.6	7.6 - 22	22 - 40.4	40.4 - 100	-

The CORSIKA simulation stops just above the snow on the ground. Propagation of secondary particles through the snow and ice in the tanks is simulated with a software package called Geant4. Geant4 supports the interaction of photon, electron, positron, muon, tau, neutrino, all hadrons, and light ions and physics processes involved during their propagation through snow and ice. The main physical processes in Geant4 are itemized below:

- Photons: The processes that Geant4 supports for photons are the photoelectric effect, Rayleigh scattering, Compton scattering, and pair production.

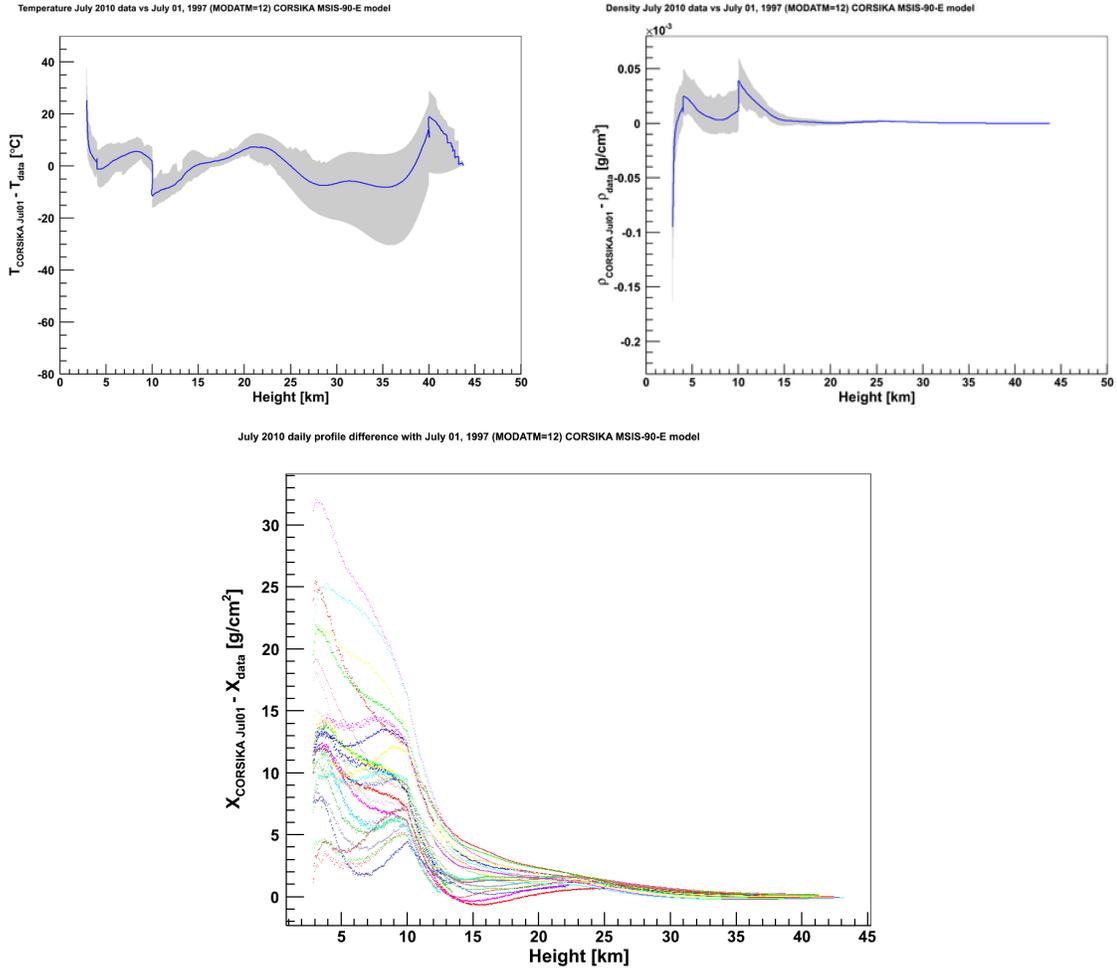


Figure 4.1: Difference between the July 2010 South Pole atmosphere, data collected by the AMRC balloon and the NASA AIRS satellite, and the July 1997 atmosphere (ATMOD12) used in CORSIKA. The top plots show the temperature and air density difference. There is unphysical behavior seen at 4 km, 10 km, and 40 km. The bottom plot shows the difference in atmospheric depth. Source of these figures is from an IceCube cosmic ray working group internal study performed by Sam DeRidder.

- Electrons and positrons: The processes that Geant4 supports for electrons and positrons are Coulomb scattering, ionization, Bremsstrahlung, and annihilation.
- Muons: The processes that Geant4 supports for muons are Coulomb scattering, Bremsstrahlung, ionization, and pair production. Also muon capture for μ^- .

- Pions, kaons, protons, and anti-protons: The processes that Geant4 supports for these hadrons are Coulomb scattering, Bremsstrahlung, and Ionization.
- Alpha and G4GenericIon: The processes that Geant4 supports for these particles are Coulomb scattering and ionization.
- The rest: Geant4 supports decay of all long-lived hadrons and leptons, and muon capture for muons.

Geant4 models the Cherenkov emission of particles in the detector to calculate the amount of light deposited in the tanks. The signal generated in the electronics due to this light is simulated using modules written by the IceCube Collaboration. Once the detector simulation is complete, Level1 and Level2 scripts are run on the simulated data. Level1 scripts perform the detector simulation and check trigger and filter conditions, exactly as is done for real data at the South Pole. Finally, Level2 scripts are run on the events after Level1 processing. This focuses on event cleaning.

4.1 GCD File

GCD stands for Geometry, Calibration, and Detector Status. The geometry part of the GCD file has information regarding the DOMs position throughout the detector. The position of a tank is defined as the midpoint of its two DOMs positions. This part of the GCD file also has information about tanks' height, radius, type of material used for inner lining, and snow height on top of them. The calibration part of the GCD file has all parameters and values after running DOMCal, VEMCal, and RAPCal. Refer to [87] for detailed information of parameters and values after calibrations. The detector status part of the GCD file has information of good and bad DOMs.

Snow height on top of each of the IceTop tanks is measured twice per year. These snow heights are interpolated to get the snow height for the data taking period. We used measured snow height from October 2016 for simulation. An important part of this analysis is that snow heights used in simulation is approximately equal to the average snow height for the data taking period, as shown in Fig 4.2.

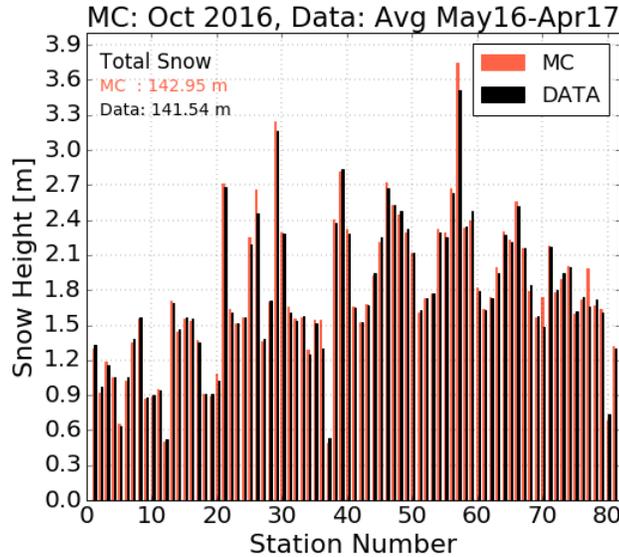


Figure 4.2: Height of snow accumulated on top of each IceTop stations in 2016. The snow heights used in simulation is from October 2016 measurement, and snow height for data is the average snow height for 2016 data taking period. The difference in the total snow on top of all stations between simulation and data is 1.41 m, which is small given that the total snow height is ~ 142 m spread over an area of 1 km^2 .

4.2 Level1

Level1 is a collection of software scripts that includes detector simulation, and trigger and filter condition check. The detector simulation is done in two steps. The first step is to read CORSIKA files and simulate the IceTop response. Geant4 is used for this step of the simulation. The second part involves reading photon-propagated files and simulating the PMT response and the DOM electronics. The detector simulation forms pulses for an event. Then the trigger condition is checked using these pulses. Modules to check trigger and filter conditions were written by the IceCube Collaboration.

To start a detector simulation, CORSIKA showers of proton, helium, oxygen, and iron primaries mentioned in table 4.1 are dropped on top of the IceTop detector model within certain re-sampling radii. If primary particles have energy from 10 TeV

to 100 TeV, showers are dropped within a radius of 200 m. If the energy of primary particles is from 100 TeV to 1 PeV, they are dropped within a radius of 400 m. If the energy of primary particles is from 1 PeV to 10 PeV, they are dropped within a radius of 600 m. If the energy of primary particles is from 10 PeV to 25 PeV, they are dropped within a radius of 800 m. Table 4.3 shows the re-sampling radii of the primary particle for various energy range. Also, the re-sampling radius superimposed on IceTop geometry is shown in Fig 4.4. These radii are selected in such a way that no simulated showers in a given energy bin that fell outside of the corresponding radius passed the new IceTop volume trigger implemented for this analysis. The true core position of simulated showers that pass the trigger condition is shown in Fig 4.3.

Once the detector simulation is done, the simulated waveforms are calibrated and the pulse is extracted. Trigger and filter conditions as described in Chapter 3 are checked using hits of an event. Any event that does not pass a trigger or a filter is removed to save disk space. If any IceTop trigger condition is passed, all information related to that event is saved. These events are now ready for Level2 processing.

Table 4.3: Resampling radius based on energy bin.

$\log_{10}[E/\text{GeV}]$	Resampling radius [m]
4.0 - 5.0	200
5.0 - 6.0	400
6.0 - 7.0	600
7.0 - 7.4	800

4.3 Level2

Level2 processing is done on events that pass any IceTop trigger and filter. Level2 focuses on cleaning events. Cleaning is done in multiple steps. It is done first with the ‘HLCClusterCleaning’ algorithm and eventually with the ‘SeededRTCleaning’ algorithm. To start cleaning, all hits of an event are sorted by time. If consecutive hits

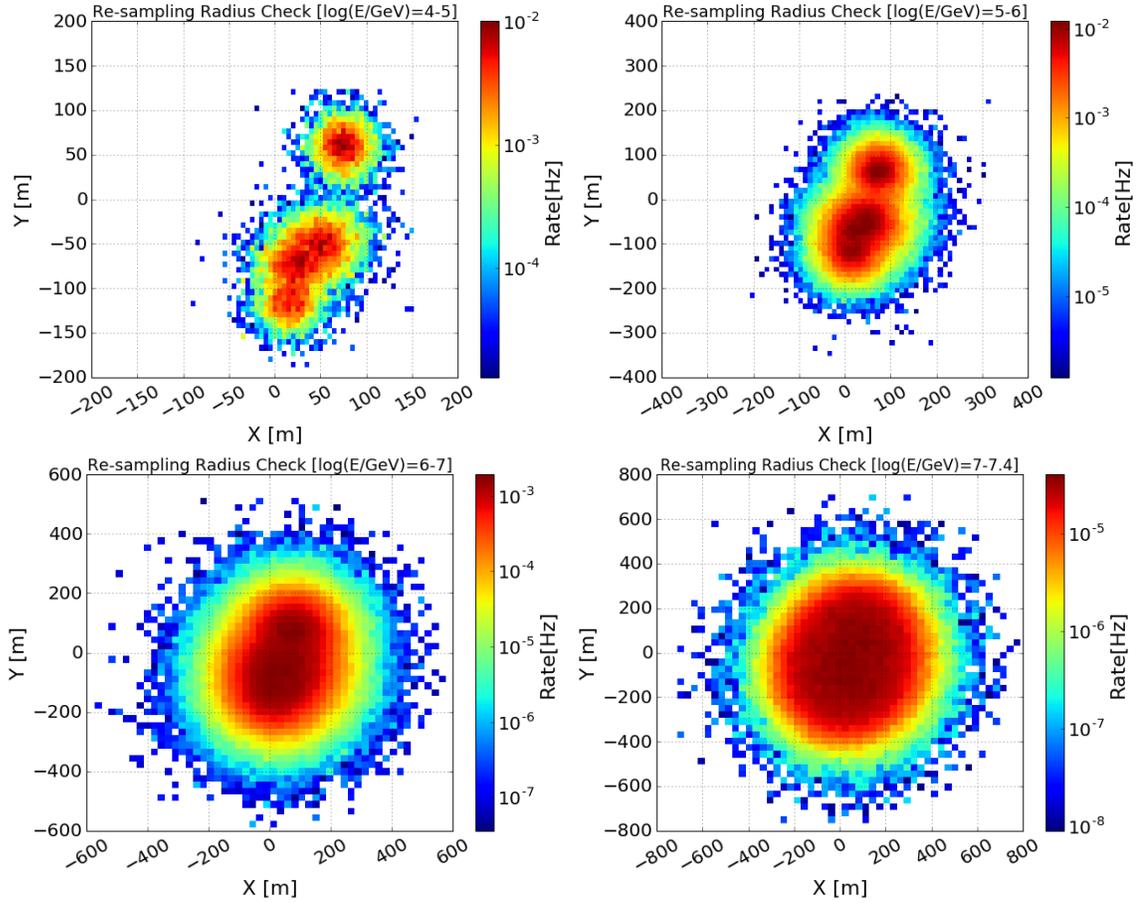


Figure 4.3: The true core position of simulated showers that pass IceTop volume trigger. Showers are contained within 200 m, 400 m, 600 m, and 800 m depending on their primary energy.

are within 200ns, they are kept. So, the cleaned event will contain a list of hits that lie within 200ns from the reset nearest hit in time.

Before ‘SeededRTCleaning’ is performed, we check for filter conditions using ‘cluster cleaned’ pulses. Events that fail IceTop_InFill_STA2 filter are removed from the analysis. To begin the process a list of few hits are kept and used as a seed. We then loop over all hits. If a hit is within R and T range, it is appended to the list. In this analysis, 180 m for R and 450 ns for T is used. The list can grow after each loop. This extended list of hits is used as a new seed to append further outlying hits within the R and T range. The iteration stops when there are no more hits to append. There are

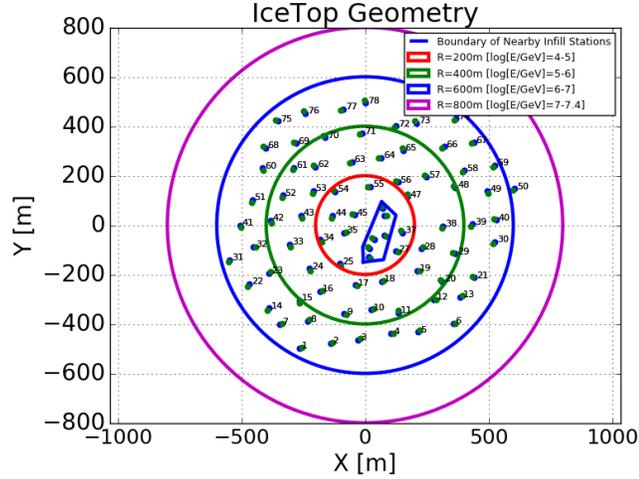


Figure 4.4: IceTop geometry with nearby InFill area inside the irregular polygon at the center. Re-sampling area for various energy bins is shown by circles. Area increases as energy of the shower increases. Radial distance increases by 200 m for one decade energy increase. Showers are re-sampled within 200 m for energy 10 TeV to 100 TeV, 400 m for energy 100 TeV to 1 PeV and so on.

various options on what to use for the initial seed. In this analysis, ‘HLCcore’ option is used as a seed, which uses all the HLC hits. An internal RT cleaning is run on these HLC hits. If a HLC hit does not have at least two hits in its R and T range, then it is removed from the seed. Removed hits can be added later during ‘SeededRTCleaning’ process if they lie within R and T range from one of the selected hits.

The ‘SeededRTCleaning’ algorithm selects pulses that belong to the event and gets rid of pulses from background hits. A check for IceTop_InFill_STA2 filter condition is re-run using these selected pulses on event-by-event basis. Events that do not pass the filter condition are removed. These pulses are used for reconstruction of the air shower’s core position, direction, and energy. In the standard analysis, a maximum likelihood method implemented by a software project called ‘Laputop’ is used for reconstruction. In this analysis, Random forest regression, discussed in more detail in Chapter 5, is used for all reconstruction.

4.4 Composition Model

Based on the element-wise flux measured by various experiments, many composition models are proposed in the cosmic ray community. These models attempt to match the real element-wise composition spectrum. For a composition model, a modified version of H4a [62] is used in this analysis. In this model, the all-particle spectrum is divided into three populations consisting of five elemental groups and is given by

$$\phi_i(E) = \sum_{j=1}^3 \alpha_{i,j} E^{-\gamma_{i,j}} e^{[-\frac{E}{z_i R_{c,j}}]} \quad (4.2)$$

where ϕ_i is a flux, γ_i is an integral spectral index, i is an index representing five elemental groups, i.e H, He, CNO, Mg-Si, and Fe. The all particle spectrum is the sum of individual flux of these groups. In this analysis, simulations for H, He, O, and Fe are generated. The oxygen simulation represents CNO and Mg-Si elemental group and is denoted by O*. The three populations are represented by j and is denoted by number 1, 2, and 3. R_c is the critical rigidity of the population. See table 4.4 for values of these parameters. The last two lines of Table 4.4 give parameters for elements from extragalactic sources and are added.¹ Population 3 parameterizes the flux of ultra high energy cosmic rays that has little to no effect in the energy range of interest.

Fig 4.5 shows the modified H4a composition model given by Eq 4.4 and parameters from Table 4.4. The left plot of the figure shows the individual flux of each elemental group scaled by $E^{2.65}$ and their sum. Ideally, the summed total flux should approximate the cosmic ray flux from experimental data. Legend O* is the combined flux of CNO and Mg-Si elemental group. The right-hand side of the figure shows the fractional composition of each elemental group compared to the total flux of the modified H4a composition model. The composition fraction of proton and helium is

¹ The H4a composition model is modified by adding contribution from both fluxes in population 3 of Table 4.4, originally intended to use flux from ‘proton only’. Since population 3 accounts for ultra high energy cosmic ray (\sim EeV) flux from extra-galactic sources, this addition does not affect the energy range of interest (TeV-PeV) in this analysis.

Table 4.4: Parameters for H4a composition model.

Rc	gamma	p	He	CNO	Mg-Si	Fe
γ for Pop.1	-	1.66	1.58	1.63	1.67	1.63
Pop.1: 4 PV	see line 1	7860	3550	2200	1430	2120
Pop.2: 30 PV	1.4	20	20	13.4	13.4	13.4
Pop.3 (mixed): 2 EV	1.4	1.7	1.7	1.14	1.14	1.14
(proton only): 60 EV	1.6	200	0	0	0	0

higher in the region of interest (100 TeV - 2 PeV) and decreases as the energy increases.

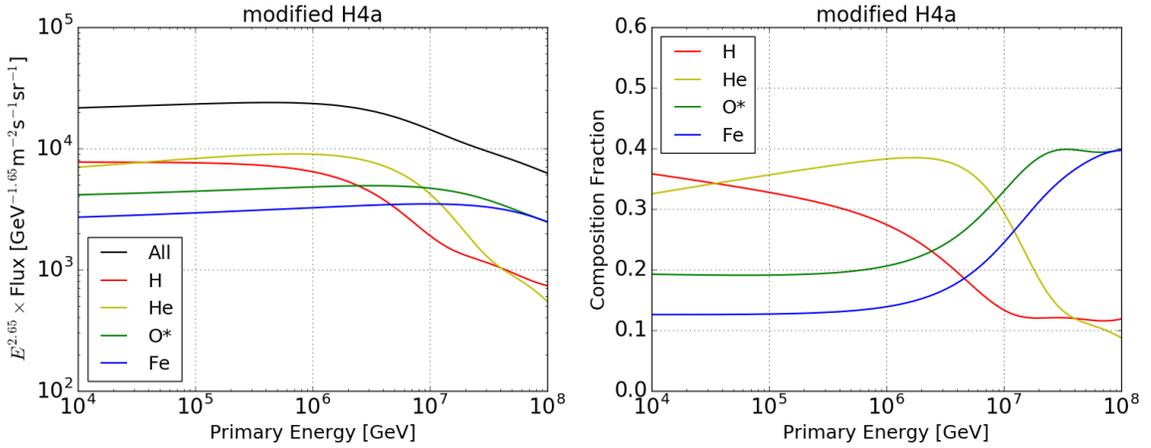


Figure 4.5: Left: modified H4a composition model where the individual flux of proton, helium, oxygen, iron, and total is scaled by $E^{2.65}$. Right: Fractional composition of modified H4a model in five elemental groups. CNO+Mg-Si is used in the analysis and is represented by O*.

In this analysis, along with modified H4a composition model, other competitive models have been used. They are GST [113], GSF [114], and a modified version of Polygonato [115]. The parameters for the modified Polygonato composition model is shown in Appendix B. These models are used to estimate the systematic uncertainty due to the assumed composition. Fig. 4.6 shows the comparison of the composition

fraction of proton, helium, oxygen, and iron. The composition fraction of each elemental group varies significantly between models.

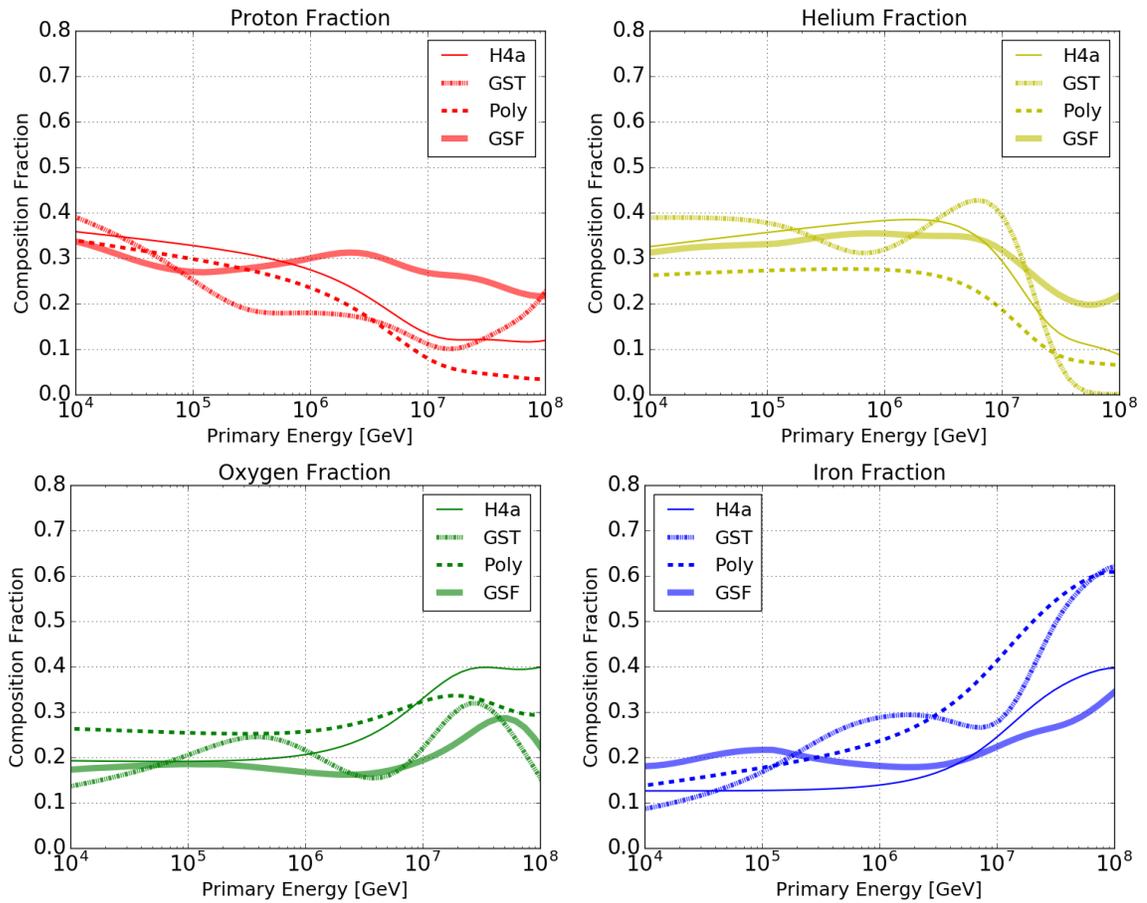


Figure 4.6: Composition fraction of commonly used 4 different composition models.

The modified H4a composition model is referred to as H4a throughout the thesis unless otherwise stated.

4.5 Weighting the Spectrum with the Composition Model

Once the Level2 simulation is completed, the final reconstruction of core position, zenith angle, and energy, and the calculation of detector efficiency and cosmic ray flux is performed. Flux calculation typically begins by calculating the rate ($R = N/\Delta t$) which is duration independent. In order for simulated data to

have a similar rate as real data, it is necessary to reweight the simulation. Simulated showers are generated with an E^{-1} spectrum and re-sampled many times. The rate of these events for given energy does not reflect the real energy spectrum. An artificial frequency for each simulated events has to be given so that the rate of simulated data approximates the rate of experimental data. The weight of each event is calculated based on the particle type and composition model. Each event is assigned a weight which will add up to give an elemental flux of its corresponding composition model. For this analysis, the composition models used to calculate the weight of each event are H4a, GSF, GST, and modified Polygonato model.

To calculate the weight, a total flux from a model and fluence from the generated simulation is required. A differential cosmic ray flux is given by

$$\left(\frac{d\Phi}{dE}\right)_{\text{expected}} = \frac{dN}{dt dA d\Omega dE}$$

where dN is the number of events in an energy bin between E and $E+dE$. dt is the duration of data taking, $d\Omega$ is the solid angle, and dA is the area over which these events landed. $(\frac{d\Phi}{dE})_{\text{expected}}$ comes from the assumed composition model.

The basic information of all events, including those that do not trigger IceTop, is saved separately for the purpose of counting the number of events and their distribution in zenith angle and energy. For example, if there are $dN_{\text{generated}}$ number of events generated in simulation over an area dA and solid angle $d\Omega$. The differential fluence for simulation is given by

$$\left(\frac{d\Phi}{dE}\right)_{\text{generated}} = \frac{dN_{\text{generated}}}{dA d\Omega dE}$$

Then the weight of events in unit of s^{-1} with energy E is given by

$$w = \frac{(\frac{d\Phi}{dE})_{\text{expected}}}{(\frac{d\Phi}{dE})_{\text{generated}}}$$

4.6 Data-Simulation Comparison

To check the quality of simulation, pulses generated after Level2 are compared with pulses from experimental data on a tank-by-tank basis for all 162 tanks. A

charge histogram of pulses after ‘SeededRTCleaning’ is used for both simulation and experimental data. A comparison is performed at various stages. If the simulation is a good representation of the data, it should behave like data at every stage.

Fig 4.7, for tank 81A, is a representative plot showing a charge histogram in arbitrary units. The figure shows good agreement between data and Monte Carlo, including detailed features of the distribution. The ratio between these two histograms is around 1. Similar agreement is found for all tanks. This simulation can therefore be implemented to predict the necessary quantities on data with reasonable confidence.

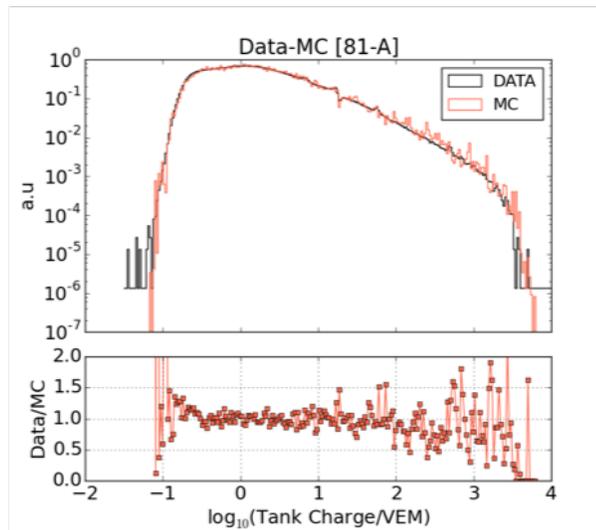


Figure 4.7: Charge histogram of tank 81A compared between data and simulation.

After tank-by-tank comparison, histograms for variables on an event-by-event basis are compared to verify the simulation. Histograms of the number of stations, the charge in each tank, the total charge of an event, the shower core position, and position of the tank that has been hit are shown in Chapter 5.

4.7 Sybill2.1 Vs QGSJETII-04

The study of the low energy spectrum is done using Sibyll2.1 as the hadronic interaction model. There are other commonly used hadronic interaction models,

like QGSJet and EPOS-LHC. As a parallel study, we generated a small number of simulations using QGSJetII-04 hadronic interaction model to compare the final result between Sibyll and QGSJet.

To look for the difference between Sibyll2.1 and QGSJetII-04, the first interaction height and slant depth at shower maximum (χ_{max}) for proton, helium, oxygen, and iron showers are compared. The top plot of Fig 4.8 shows the comparison of first interaction height histogram for four different types of primary particles between Sibyll and QGSJet. The difference in mean height between these two models is less than 0.35 km at an altitude more than 20 km. Similarly, the bottom plot of Fig 4.8 shows the comparison of shower maximum for four different types of primary particles between Sibyll and QGSJet. The mean shower maximum from the distribution is relatively close to each other. For all particles, the difference of the mean is less than 18 g/cm² when the shower maximum is more than 480 g/cm². An overview of the differences between Sibyll and QGSJet hadronic interaction models is given in [116, 117]. These differences may result in difference between cosmic ray energy spectra.

The all-particle cosmic ray energy spectra is produced separately using simulations with Sibyll 2.1 and QGSJetII-04 as hadronic interaction models and results are given in Chapter 6.

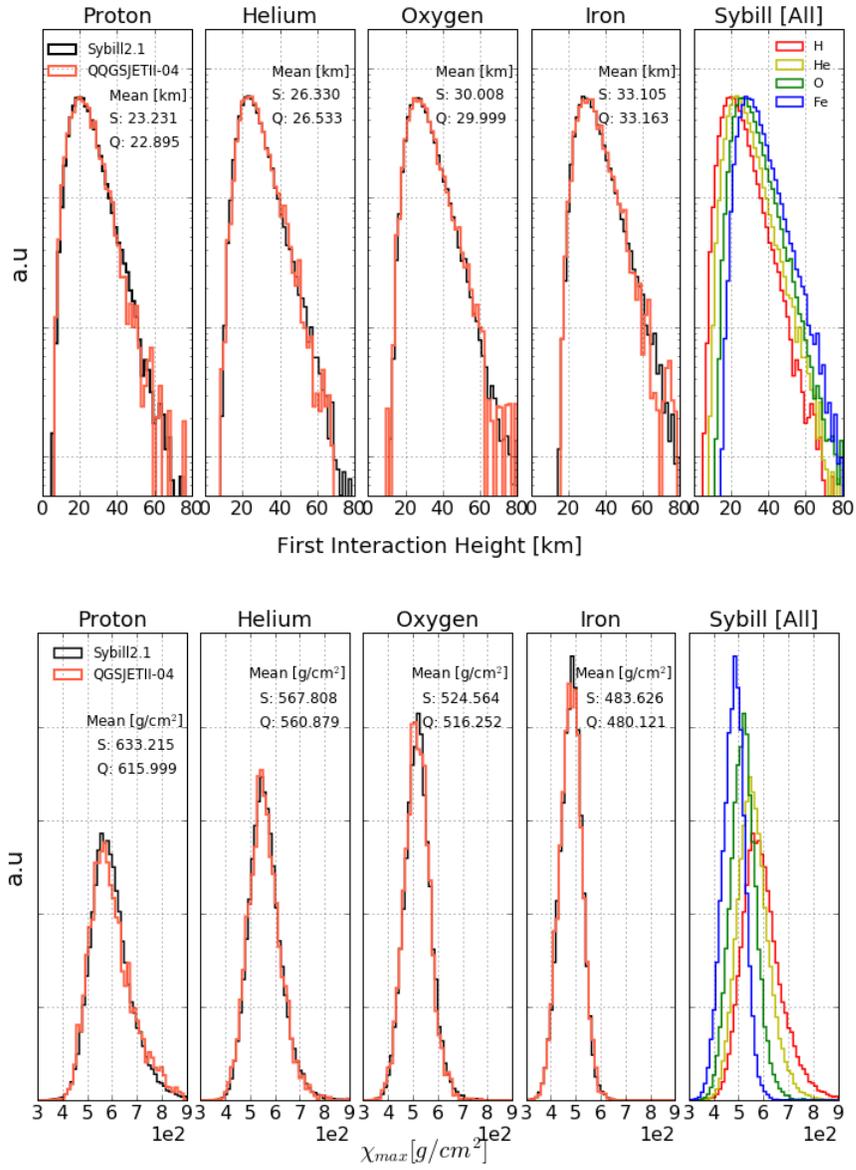


Figure 4.8: Comparing quantities between Sibyll2.1 and QGSJetII-04. Top: First Interaction height between Sibyll2.1 and QGSJetII-04 for four elements. Bottom: Shower maximum between Sibyll2.1 and QGSJetII-04 for four elements.

Chapter 5

ICETOP AIR SHOWER RECONSTRUCTION

The cosmic ray flux approximately follows a power law. As the energy increases, the flux decreases. The approximate integral flux of cosmic rays with energy greater than E is given by $I(> E) \approx 1 \times 10^4 (E/\text{GeV})^{-1.7} \frac{\text{nucleons}}{\text{m}^2\text{ssr}}$ [1]. If a detector has an acceptance of $0.1 \text{ m}^2\text{sr}$, then it can receive approximately 100 showers with energy higher than 100 TeV in 1 year. Similarly, for showers higher than 10 PeV, it can detect approximately one event every 25 years. The rate of events in the energy range of interest, $E \geq 250 \text{ TeV}$, is less than $10 \text{ m}^{-2}\text{sr}^{-1}\text{yr}^{-1}$. Hence, an indirect measurement with a large extensive air shower detector on the ground is needed to detect many events in a reasonable time. This ability to detect high energy events comes with a price, i.e., not knowing their primary energy and direction. These unknown quantities have to be reconstructed from the known and detected quantities.

IceTop is a ground-based detector array spread over an area of 1 km^2 . The information gathered by IceTop is the charge and time of hit for each tank. Besides charge and time of the hit, the tank position is also known. Therefore, the hit time, charge, and position of each tank are used to reconstruct the necessary information of air showers.

Previous studies [98, 99, 100, 101] used 5 or more stations to reconstruct the air shower's core position, direction and energy using the software project called 'Laputop'. Laputop fits a lateral distribution function (LDF) on known quantities and maximizes the likelihood. As a result, core position, direction, and energy proxy are produced as parameters for each event. However, Laputop has limitations on reconstructing events with fewer than five stations hit. For the analysis in this dissertation, we develop a machine learning approach to event reconstruction.

This chapter describes the machine learning technique and its implementation to reconstruct two-station events. The features that go into machine learning for reconstruction are discussed. Then the resolution for core, zenith, and energy that determine the goodness of reconstruction are shown. Simulation is a vital component of reconstruction. So, comparisons between data and simulation are also shown in this chapter.

5.1 Machine Learning

Machine learning is a set of computer algorithm designed to train on a provided data set and develop a model of the data. The algorithm trains itself and uses the generated model to predict. In this analysis, supervised learning is used to train a model to predict core position, direction, and energy of each air shower. For training and prediction, a simulated data set with known core positions, direction and energy for each event is used. The same set of features, like position of hit tanks, charge on hit tanks, total charge deposited in all hit tanks etc., used in training are given from experimental data as an input to the trained model built with simulation. Based on these input features, the model predicts core position, direction, and energy for events in the experiment.

Random forest regression is the algorithm used here for reconstruction¹ of the core position, zenith angle, and energy. A random forest is a ‘forest’ made up of many decision ‘trees’. The following subsection describes the working mechanism of random forest regression algorithm.

5.1.1 Random Forest Regression

The formation of a decision tree involves recursively splitting data into two parts as shown in Fig 5.1. The split is done based on a feature that minimizes the *Residual*

¹ “Prediction” is a commonly used term in machine learning for the reconstructed value of quantity in data. For example, “predicted energy” is the result of applying the trained algorithm to data, which is more commonly referred to as “reconstructed energy”.

Sum of Squares (RSS) at each node. To decide on which feature minimizes the RSS, data are divided into two parts by various values of each used features. The RSSs are calculated on both parts and are added for all values and all used features. Then all summed RSSs are compared. The final split of data at a node is done by a feature at a value that produces the minimum RSS. As an example, assume there are n_1 and n_2 events after splitting data by a feature at a value while predicting energy. Assume \hat{E}_1 is the mean true energy of n_1 events and \hat{E}_2 is that for n_2 events. The RSS is given by

$$RSS = \sum_{i=1}^{n_1} (E_i - \hat{E}_1)^2 + \sum_{i=1}^{n_2} (E_i - \hat{E}_2)^2 \quad (5.1)$$

where E_i is true energy of event i that ends up in one of the two groups. Once the data is split into two parts, the process is repeated by further splitting the split data. Again a split is done based on a feature and its value that minimizes RSS. A detailed description of random forest regression is found in [118].

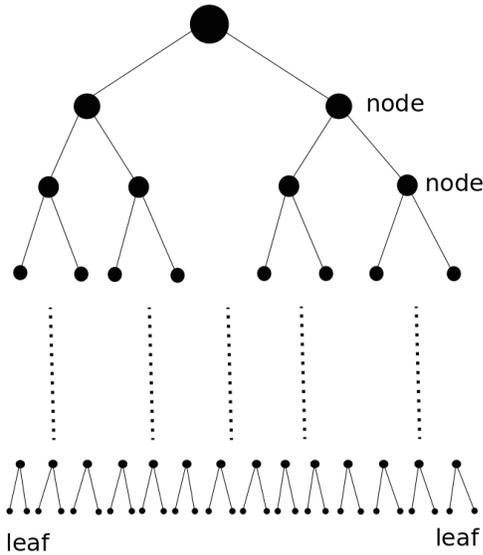


Figure 5.1: Schematic of a decision tree. Data is split into two parts by a feature that minimizes Residual Sum of Squares (RSS) at each node.

Branching of data by splitting it multiple times leads to a tree-like structure. The end of a tree is called a leaf. The branching process continues until a stopping

criterion is reached. There are two commonly used stopping criteria. The first is a maximum number of branches in a tree which is commonly known as the depth of a tree. The second is the minimum number of events at a leaf of a tree. If one of these criteria is fulfilled then the splitting process is complete. Each leaf of a tree uses the mean of the true value from events that land in that leaf and assign that mean value as its prediction. This tree-like structure formed by splitting data is called a decision tree.

A forest has many decision trees. A variation among trees comes from using randomly selected subsets of features and data. A decision tree is grown by considering randomly selected features on each split. If there are f total features and only m are used in each split then usually $m \approx \sqrt{f}$ is used. Not all data are used to grow a tree. Data are randomly selected by a bootstrapping method. The number of bootstrapped datasets is equal to the number of trees in a forest. A random forest is generated by tuning several features in each split from randomly selected bootstrapped data. During training the algorithm randomly selects data used to grow a tree so that each decision tree is slightly different. Combining the predictions from each tree reduces the variance of the prediction [119]. Finally, the unknown quantity is predicted by knowing which leaf the event lands on in each tree and taking an average of their predicted values.

5.2 Machine Learning Features

Various features of an air shower are used to predict (reconstruct) core position, zenith angle, and energy. Not all input features are equally important. Some features are good at predicting an unknown quantity while others do not help. A feature that is used more times to split data has a higher feature importance. Random forest regression uses gini-index to calculate feature importance of each features used in training. A description of gini-index and other techniques to calculate feature importance is discussed in Chapter 8 of [118]. In this section, a detailed explanation of features used in this analysis for reconstruction of air showers is given.

5.2.1 Shower Core with Center of Gravity (COG)

Shower core is the position where the shower axis hits the ground. It is calculated as the weighted sum of positions of tanks that have been hit. The weight used is the square root of the charge on an individual tank. The x -coordinate of the shower COG is given by

$$\text{COGX} = \frac{\sum_i \sqrt{Q_i} x_i}{\sum_i \sqrt{Q_i}} \quad (5.2)$$

where i runs over tanks that have been hit, Q_i is the charge on tank i , and x_i is the x -coordinate of tank i . Similarly, COGY is calculated using y -coordinate of hit tanks. It is shown in Fig 5.8 that COG has high feature importance while predicting the shower's core position.

5.2.2 Direction of Plane Shower Front (θ_{plane} , ϕ_{plane})

A plane shower front traveling perpendicular to the shower axis is a good approximation for events with a small number of hit tanks. Assuming a plane shower front hits the ground at shower core at time t_0 and hits tank i at position (x_i, y_i) at time t_i . The quantities of interest are zenith angle (θ) and azimuth angle (ϕ) of the primary particle that caused the air shower. The initial reconstruction is done by minimizing chi-square given by

$$\chi^2(n_x, n_y, t_0) = \sum_i \frac{(t_i - t_0 + \frac{n_x x_i + n_y y_i}{c})^2}{\sigma_i^2} \quad (5.3)$$

where t_i is the measured signal time, $n_x = \sin \theta \cos \phi$, and $n_y = \sin \theta \sin \phi$. The sum runs over all hit tanks. The time uncertainty σ_i is equal to 5 ns for all tanks [87]. The plane fit direction is $\theta = \cos^{-1}(-\sqrt{1 - n_x^2 - n_y^2})$ and $\phi = \tan^{-1}(\frac{n_y}{n_x})$. It is shown in Fig 5.8 that θ_{plane} has a high feature importance and predicts shower's zenith angle.

5.2.3 Average Z in Shower Coordinates (ZSC_{avg})

ZSC_{avg} is the average distance of hit tanks from the plane shower front when the core hits the ground. ZSC_{avg} is higher for inclined showers and approximately zero

for vertical showers. It is given by

$$Z_{\text{avg}} = \frac{\sum_{i=1}^n |z_i|}{n} \quad (5.4)$$

where i runs over n hit tanks and z is the position of a tank in the shower coordinate system. It is shown in Fig 5.8 that Z_{avg} has reasonable feature importance while predicting shower's zenith angle.

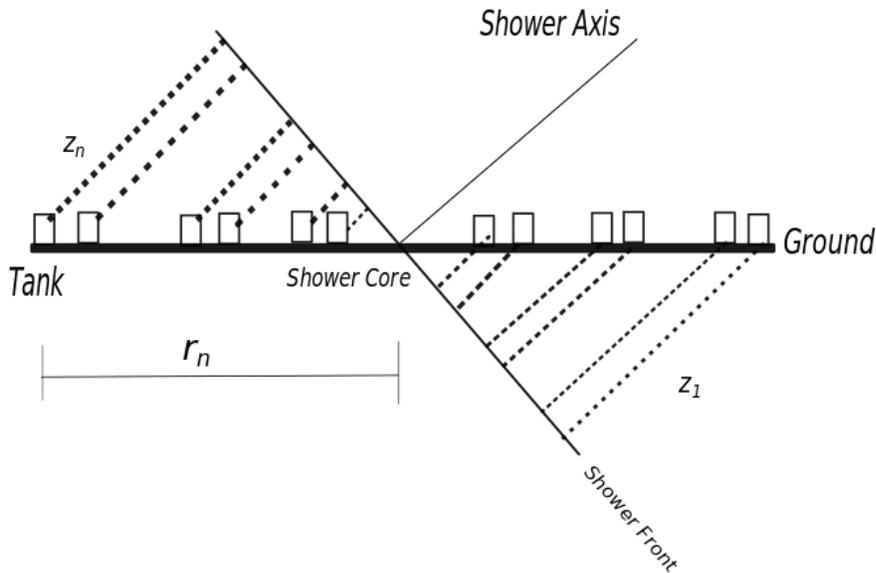


Figure 5.2: Schematic of a plane shower front hitting IceTop tanks. Z_{avg} is the average of z in the shower coordinate system. r is the distance from the shower core to a hit tank.

5.2.4 Tank Distance from Shower Core (hitR)

hitR is the distance from the shower core of each tank that has been hit. In Fig 5.2 r_1 is the distance from the shower core of the first tank that has been hit. hitR consists of an array of distances of all tanks that are hit. The distance of a hit tank can be either from the ShowerCOG core position or the predicted core position. For energy prediction, hitR is calculated from the predicted shower core. It is shown in Fig 5.8 that hitR has a high feature importance while predicting shower's primary energy.

5.2.5 Other

Other features that go into the machine learning are hitX, hitY, and the total number of hit stations (N_{sta}) in each event. A sum of charges (Q_{tot}) from all hit tanks and the sum of the two highest charges (Q_{sum2}) of an event are also used. To explain some features that go into the random forest regression, suppose a shower hits n tanks and deposits some energy in each tank. We form a list of charges on these tanks and sort them in descending order. The list is denoted by hitQ. We also arrange the position of tanks based on their corresponding charge, largest to smallest. This list is denoted by hitX and hitY representing the x and y coordinate, respectively. The minimum value that n can have is four and it can increase up to 162 as the energy of primary particle increases. Hence, the minimum number of charge and position information of an event is four. For the energy region of interest, information from the first 35 hits is enough to reconstruct shower core position, direction, and energy with almost 100% feature importance. The cumulative feature importance as a function of a number of hits is shown in Fig 5.3.

The list of charges (hitQ) is stored in descending order. The list of tank positions (hitX, hitY) is arranged based on the charge on hit tanks. If the number of tanks (n) that have been hit is less than 35, then the remaining $35-n$ slots of the list are filled with 0 for hitQ, hitX, and hitY. hitT denotes the list of times at which tanks have been hit for each event. The time of the first hit of an event is subtracted from all hits such that the time used is with respect to the first hit. The time information of an event is arranged in ascending order. If an event has fewer than 35 hits, then the remaining slots are filled with the last relative time.

The random forest regression becomes computationally expensive as the number of features increases. Therefore, the number of items per event in each list is truncated to 35 from 162. Fig. 5.3 shows the cumulative distribution function of feature importance for predicting shower core and zenith. It shows that a feature with the highest importance saturates before reaching 35 number of hits. If the number of hits is less than 35, those spaces are filled with either zero or the last entry. The highest

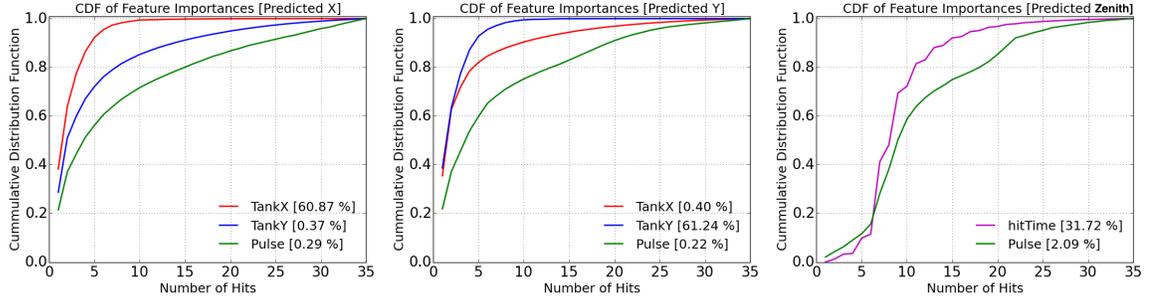


Figure 5.3: Feature Importance of 2D features to predict air shower’s core position (X and Y) and zenith angle (θ). The maximum feature importance saturates before reaching 35, so it is a good idea to use information from first 35 tanks instead of all 162 tanks.

value, or a mean of a feature, can also be used to fill, but this change has a negligible effect on resolution of core position and zenith angle.

5.3 Reconstruction using Spark

As the number of trees becomes larger with increasing depth, a single node/computer can have difficulty in computing random forest regression due to lack of memory. Also, the memory and time required to complete a job increases rapidly if many parameters are used to tune the random forest regression. So it is better to distribute jobs into multiple nodes to use more CPUs and memory. This parallel processing of a job is done using a software package called Apache Spark [120]. A comparison of times required to complete a job by running it in a single core and in parallel in many cores (CPUs) using spark is shown in Fig 5.4.

Random forest regression trains on many decision trees separately, so the training can be done in parallel. Spark is a fast cluster computing technology designed to run jobs in parallel. It increases the processing speed of an application by reducing the number of read-and-write operations to disk. This goal is achieved by storing the intermediate processing data in memory. The data structure used in Spark is in a format called Resilient Distributed Datasets (RDD) [121]. RDD can be divided into logical partitions that can be used to compute in parallel. During computation, Spark

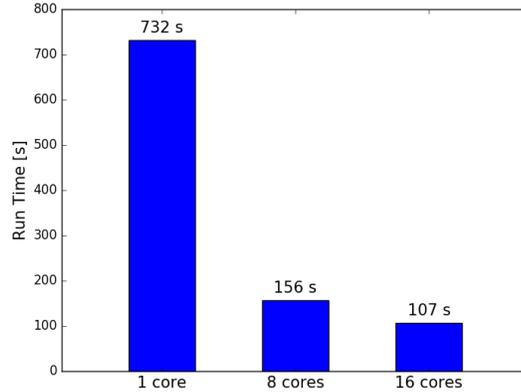


Figure 5.4: Reduce in time while running a job in multiple cores using Spark. Time required to complete an example job reduces from 732 s using 1 core to 107 s using 16 cores.

keeps track of the lineage of partitioned data so that the failed part of the computation, if any, can be recomputed.

5.3.1 Reconstruction of Core Position

Reconstruction of the core position includes reconstructing x and y coordinates of the core separately. Simulated data are randomly shuffled and divided into two halves to avoid using the same data for training and prediction. If the first half is used for training, the model it generates is used to predict the second half and vice-versa. The same technique is used for training and predicting zenith angle. Features that are used to train and predict x and y coordinate of shower's core position are:

- list of x and y position of tanks that have been hit (hitX, hitY)
- shower COG (COGX, COGY)
- list of charge on tanks that have been hit (hitQ)
- cosine of zenith calculated assuming plane shower front ($\cos \theta_{\text{plane}}$)
- $\log_{10}(\text{Number of stations})$ (INsta)

hitX, hitY, and hitQ are two-dimensional arrays with shape $N \times 35$ where N is the number of events.

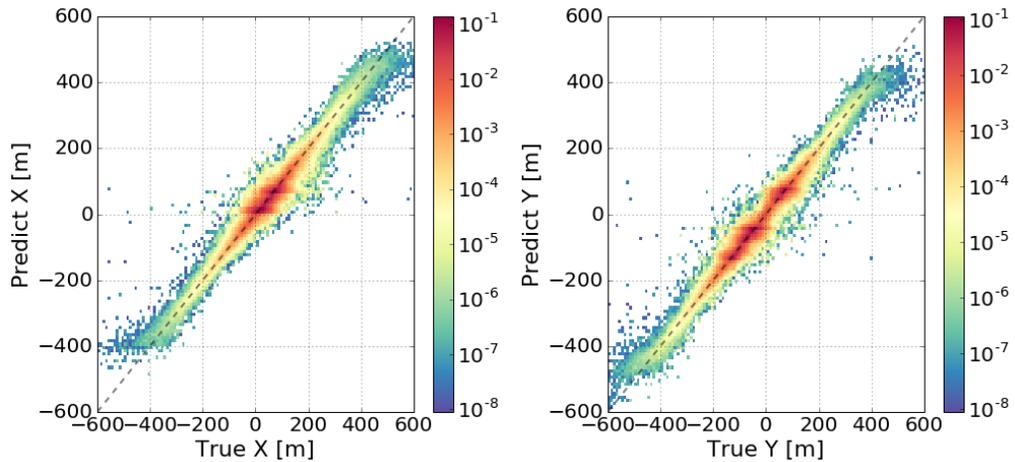


Figure 5.5: True vs reconstructed (predicted) x -coordinate and y -coordinate of core positions before final quality cuts.

Random forest regression algorithm provided in Spark package is used to reconstruct each shower’s core position. The prediction capability of machine learning depends on the input data. Quality cuts are implemented before training to remove events that will be poorly reconstructed. Removing these events improves the prediction. In the case of core position, all events that pass the IceTop_InFill_STA2 filter are used except events in which Qsum2 contains more than 95% of the total charge. Events that have majority of energy deposited in one or two tanks are called balloon events and are known to have a poor reconstruction of core position.

In an effort to find the best model, a threefold cross-validation is implemented on the input data for parameter tuning. During cross-validation, the training dataset is divided into three equal parts. The training is performed on two parts, and the model it generates is tested on the remaining part for each input parameter. Once the best parameter set is identified, the training is done on the entire training dataset using the best parameter. After running cross-validation, 1000 trees with a maximum tree

depth of 14, and minimum number of 10 events per node are found to be the optimum parameters.

5.3.2 Reconstruction of Zenith Angle

Reconstruction of zenith angle is also done using Spark. Features that are used to train and predict zenith angle are itemized below.

- List of the tank hit times (hitT)
- Shower COG (COGX, COGY)
- List of charge on hit tanks (hitQ)
- Zenith calculated assuming plane shower front (θ_{plane})
- Azimuth calculated assuming plane shower front (ϕ_{plane})
- $\log_{10}(\text{Number of stations})$ (INsta)
- Average z position of tanks that have been hit in shower coordinate system (avgZsc)

Again hitT and hitQ are the two-dimensional arrays with shape $N \times 35$ where N is the number of events. The same hitQ used while predicting core position is used here to train and predict zenith. hitT, as described in section 5.2, is arranged in ascending order and is implemented to predict zenith angle.

Random forest regression algorithm provided in Spark package is used to reconstruct each shower's zenith angle. As mentioned before, the prediction capacity of machine learning increases if events that will be poorly reconstructed are removed from input data. Therefore, the same quality cuts used to train and to predict core position are implemented. After running cross-validation for parameter tuning using Spark, the optimum parameters for predicting zenith angle are 1000 trees with a maximum tree depth of 14, and minimum number of 10 events per node. The two-dimensional histogram of true and predicted zenith is shown in Fig 5.6.

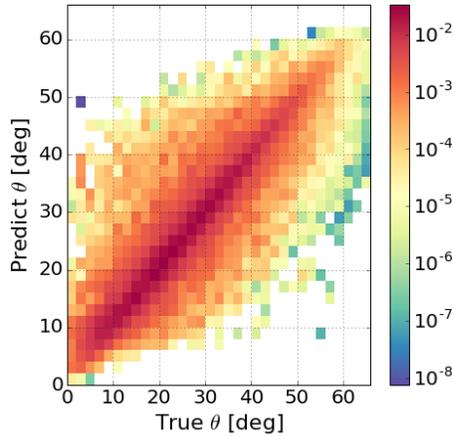


Figure 5.6: True vs reconstructed zenith angle before final quality cuts.

5.4 Reconstruction using Scikit-Learn

A random forest regression is used from a software package called Scikit-Learn [122] to predict the primary energy of each shower. The algorithm for random forest regression is the same as in Spark. The only difference is the ability of Scikit-Learn to weight an input event during training by a realistic composition model that Spark lacks. The input weight of events during training removes an energy-dependent bias on the predicted energy.

5.4.1 Reconstruction of Energy

Reconstruction of energy is done using Scikit-Learn. Features that are used to train and predict energy are itemized below.

- List of a distance of hit tanks from the predicted shower core (hitR)
- List of charge on hit tanks (hitQ)
- Cosine of zenith predicted using spark ($\cos \theta_{\text{pred}}$)
- Total charge of an event (Q_{total})

- Sum of first two highest charges of an event ($Q_{\text{sum}2}$)
- $\log_{10}(\text{Number of stations})$ (INsta)

Again hitR and hitQ are the two-dimensional arrays with shape $N \times 35$ where N is the number of events. The arrays hitQ and hitR are described in subsection 5.2.5.

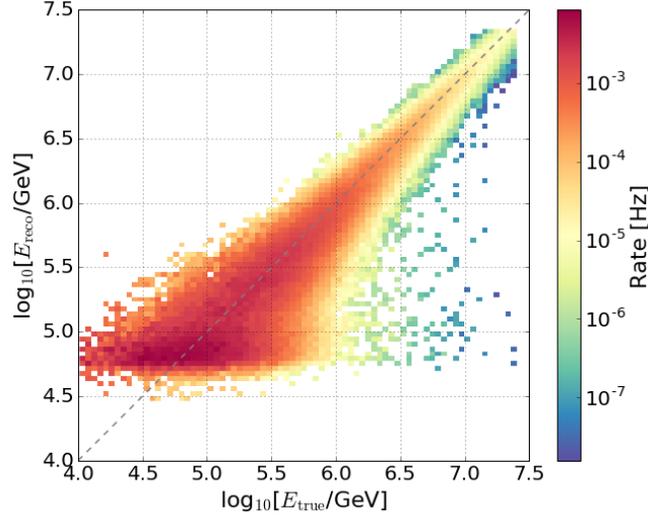


Figure 5.7: True vs reconstructed energy before final quality cuts.

A few additional quality cuts are implemented to remove events that will be poorly reconstructed. On top of what has been used for core position and zenith, events without the maximum charge (Loudest Station) on one of the twelve trigger tanks are removed. Also, events with the cosine of predicted zenith angle less than 0.8 are removed. After running cross-validation for parameter tuning, 600 trees, ten maximum depth, and ten minimum number of events per node are found to be the optimum parameters. While training, events have been weighted by the H4a composition model. The two-dimensional histogram of true and predicted energy is shown in Fig 5.7

5.5 Reconstruction Quality

Random forest regression trains itself on input features and selects the feature that helps it to predict better. Fig 5.8 shows the importance of features in percentage

to predict core position (X and Y), zenith angle, and energy. As shown in Fig 5.8,

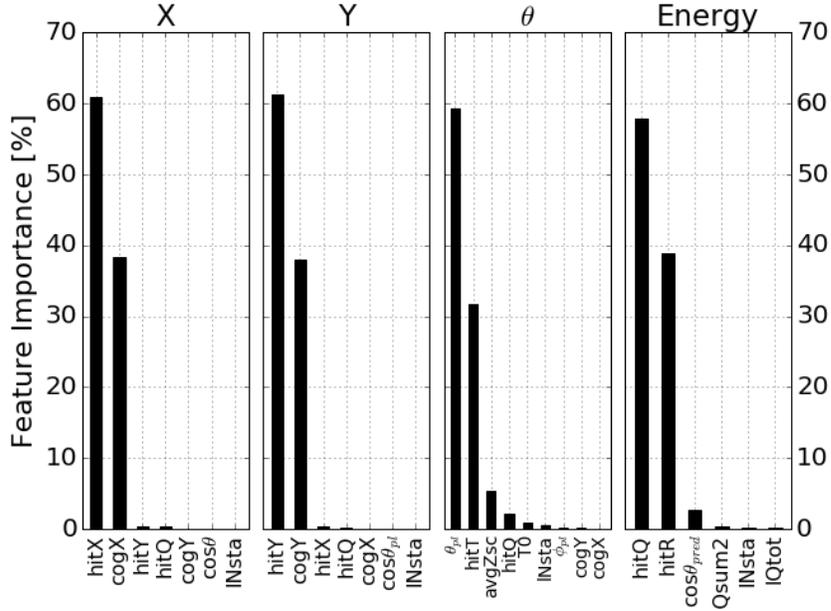


Figure 5.8: Feature importance to predict x and y coordinate of core position, zenith angle, and energy. List of coordinates of hit tanks has the highest feature importance to predict core position. The zenith angle assuming plane shower front and time of hits have a higher feature importance to predict zenith angle. List of charge on hit tanks and their distance from the shower core have a higher feature importance to predict energy.

the feature importance of the position of hit tanks and ShowerCOG is higher than the rest of the features to predict a shower’s core position. For zenith angle prediction, the angle of the plane shower front has the highest feature importance. The next important feature is the hitT. Similarly, the feature importance of hitQ and hitR is higher than the rest of the features for energy prediction.

The quality of reconstruction is judged based on how good the resolution is. The resolution is defined as one standard deviation of the distribution of the difference of true and reconstructed quantity. Fig 5.9 shows core resolution, zenith resolution, and energy resolution. The core resolution is about 16m for the lowest energy bin ($\log_{10}[E/\text{GeV}]$ 5.4 to 5.6) and improves (resolution decreases) as energy increases, the zenith resolution is about 4° for the lowest energy bin and improves as energy increases,

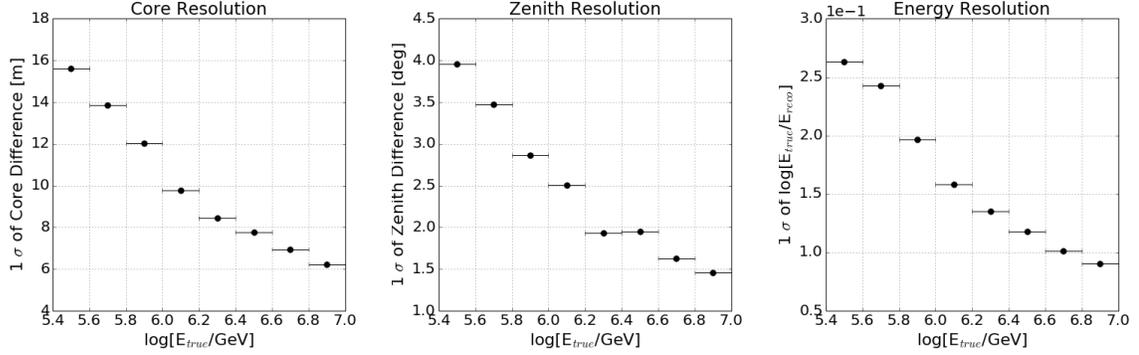


Figure 5.9: Resolution is defined as one standard deviation of true minus reconstructed distribution. Left: Core resolution in meter; Middle: zenith resolution in degree; Right: energy resolution in unit-less quantity. The difference distribution comes from the difference of $\log_{10}[E/\text{GeV}]$ value of true and reconstructed. See Table A.4 for resolution values.

and the energy resolution is about 0.26 in $\log_{10}[E/\text{GeV}]$ for the lowest energy bin and improves as energy increases. At the highest energy bin ($\log_{10}[E/\text{GeV}]$ 6.8 to 7.0), core resolution is 6.2 m for random forest regression method and ~ 7.5 m [100] for ‘Laputop’ reconstruction. Similarly, energy resolution is 0.09 for random forest regression and 0.06 [100] for Laputop. Analyses using Laputop reconstruction provide only directional resolution which can not be compared with zenith resolution.

5.6 Data and MC Comparison

A simulation is a vital tool for the analysis of low energy spectrum. Several plots are made to compare data and simulation and are shown in figures 5.10, 5.11, and 5.12.

The reconstruction method developed in this thesis uses simulated air shower data to train the machine learning models. It is therefore important that the simulation correctly reproduces low-level details of the data. Figures 4.7, 5.10, 5.11, and 5.12 show that the simulations represent the data correctly in great detail. For example, fine details that reflect the geometry of the tank locations viewed from X and Y directions are correctly represented, as is the selection effect that requires the highest charge to

be on one of the twelve triggering tanks in the middle of IceTop. The discrepancy for N_{station} more than 35 is an artifact of the analysis that does not affect the spectrum in the energy range of interest.

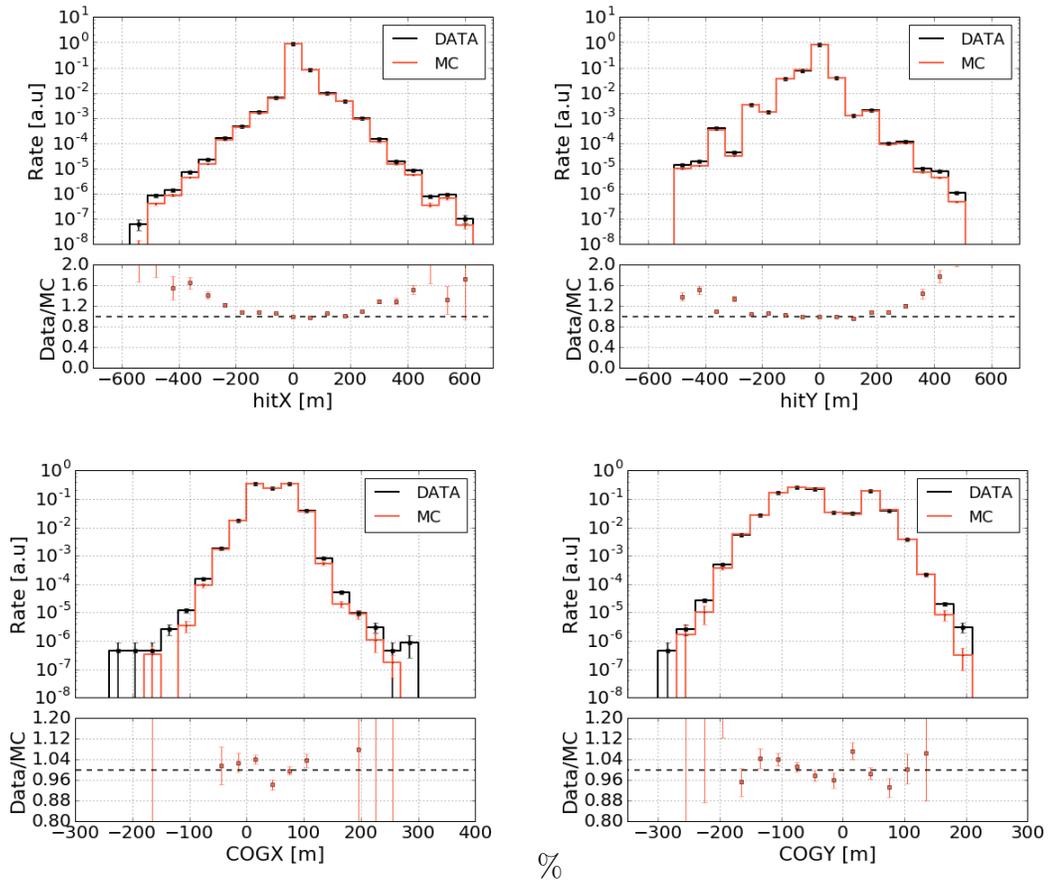


Figure 5.10: Top: Position of tanks that have been hit. Bottom: COG position of shower core.

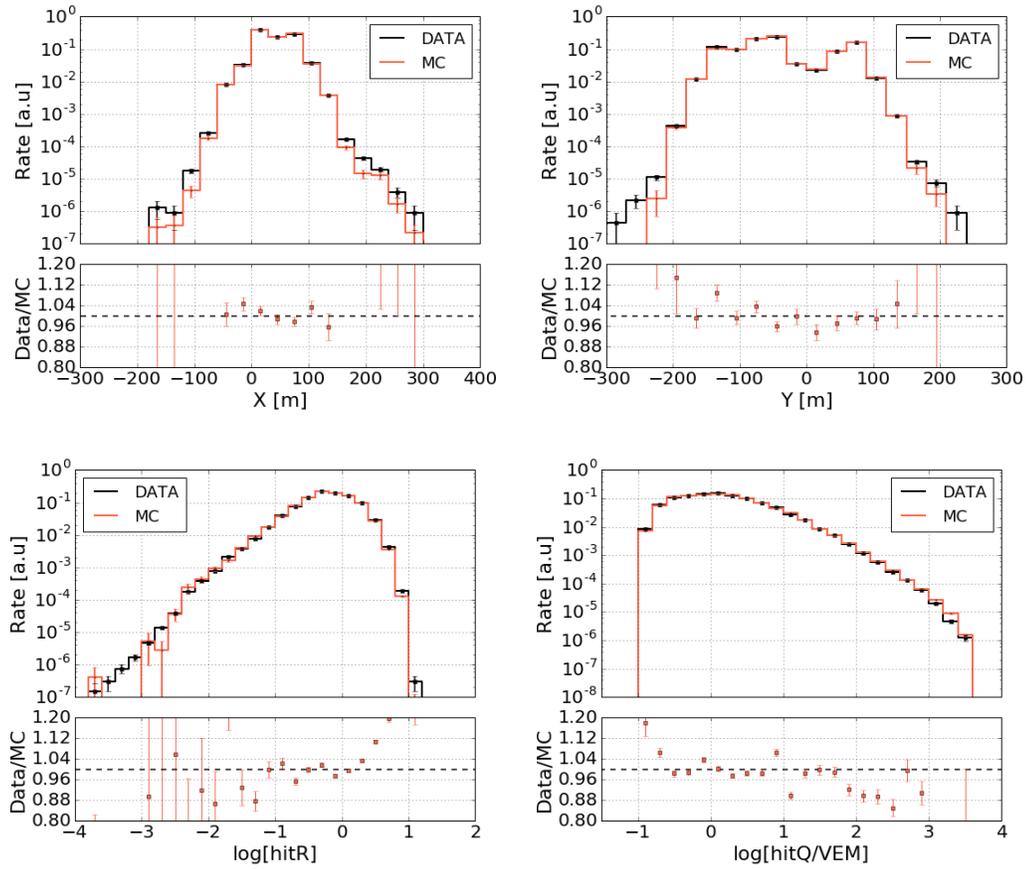


Figure 5.11: Top: Predicted X and Y coordinate of core position of showers. Bottom: Charge on each hit tanks and the distance of hit tanks from the predicted showers' core position.

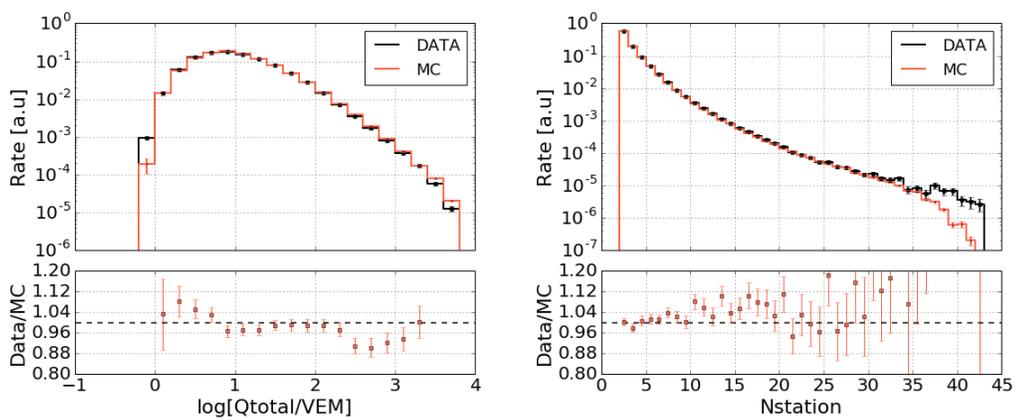


Figure 5.12: Left: Comparison between total charge histogram of events from data and simulations. Right: Comparison between total number of stations histogram from data and simulations.

Chapter 6

ICETOP LOW ENERGY SPECTRUM

The cosmic-ray flux, or energy spectrum, is the distribution of cosmic rays by energy. To measure this, a ground based detector like IceTop is required to reconstruct the properties of extensive air showers and to know the detector efficiency. Once properties such as the core position, direction, and energy are reconstructed, and the detector efficiency is known, the cosmic ray flux is calculated. The cosmic ray flux is given by

$$J(E) = \frac{1}{\epsilon(E)} \frac{dN(E)}{d\ln E d\Omega dA dt} \quad (6.1)$$

The binned flux is given by

$$J(E) = \frac{1}{\epsilon(E)} \frac{\Delta N(E)}{\Delta \ln E \Delta \Omega A_{\perp} T} \quad (6.2)$$

where $\epsilon(E)$ is the detector efficiency for events with energy E . $\Delta N(E)$ is the number of events with energy per logarithmic bin of energy, A_{\perp} is the area perpendicular to the shower's direction, $\Delta \Omega$ is the solid angle, and T is the duration of data taking. The mean area perpendicular to the direction of incidence for an observed zenith range $[\theta_1, \theta_2]$ is $A_{\perp} = A \times (\cos \theta_1 + \cos \theta_2)/2$, where A is the area of the array. The solid angle for that zenith range is $\Delta \Omega = 2\pi(\cos \theta_1 - \cos \theta_2)$. Combining these quantities together Eq 6.2 is written as

$$J(E) = \frac{\Delta N(E)}{\Delta \ln E \pi (\cos^2 \theta_1 - \cos^2 \theta_2) A_{\text{eff}} T} \quad (6.3)$$

where A_{eff} is the effective area and is given by $A \times \epsilon(E)$. The effective area for IceTop events with $\cos \theta \geq 0.9$ is shown in Fig 6.2.

This chapter is organized as follows. It begins with quality cuts that are implemented to select events. Then quantities used in the flux calculation are described.

The quantities used are the duration (T) and effective area (A_{eff}). The flux calculated with reconstructed energy has some bias due to energy bin migration. To account for this migration, DAgostini’s iterative Bayesian unfolding process is used and is explained in more detail in [123]. The basics of iterative Bayesian unfolding is described next. Before the flux is calculated, its systematic uncertainty is studied. These uncertainties are explained next. Then the result of this analysis, the cosmic ray flux with 2016 IceTop data, is given, followed by checks to justify why the specific prior (described in Section 6.4) and bin center are selected. Finally, a new analysis is performed using QGSJetII-04 as a hadronic interaction model.

6.1 Quality Cuts

To obtain the energy spectrum, only well-reconstructed events are used. Quality cuts are used to remove events with possible bad reconstruction to reduce error and to improve accuracy. The following cuts are applied to the simulated and the experimental data:

- IceTop_InFill_STA2 filter passed
- Loudest station must be inside the nearby infill boundary.
- $\cos \theta \geq 0.90$.
- Balloon cut
- High energy cut

To elaborate on event selection, the two-station filter as described in Chapter 3 has to be passed. The loudest station is the station with the highest charge tank. The highest charge tank should belong to stations 26, 36, 46, 79, 80, and 81, which are used to describe two-station trigger and filter. This cut is designed to select events with shower core near the two-station infill boundary. The loudest station cut supersedes the standard IceTop containment cut, where a shower core is inside the outermost ring of IceTop stations. Events with $\cos \theta \geq 0.9$ are used. These events have higher triggering

efficiency and are better reconstructed as compared to events with $\cos \theta < 0.9$. An event with most of its charge collected in a single tank is known to be poorly reconstructed. Therefore, these events are removed using the balloon cut. The balloon cut implements the following conditions for the largest charge (Q_{\max}) and the sum of the two largest charges (Q_{sum2}).

$$Q_{\max} \leq 0.75 \times Q_{\text{total}}$$

$$Q_{\text{sum2}} \leq 0.9 \times Q_{\text{total}}$$

Finally, the last cut is a high energy cut. The maximum energy of a simulated event is $\log_{10}[E/\text{GeV}]=7.4$. There are events with energy higher than $\log_{10}[E/\text{GeV}]=7.4$ in the experiment. These events must be removed so that data and simulation can be compared. Therefore, the following conditions are implemented:

$$N_{\text{station}} \leq 42$$

$$0.63 \text{ VEM} \leq Q_{\text{total}} \leq 10^{3.8} \text{ VEM}$$

6.2 Duration

Duration, also called live time, is calculated for each run by fitting the distribution of time differences between adjacent events. The fit function is an exponential decay function given by $I_0 \exp(-\frac{t}{\tau})$ where I_0 is the normalization constant, t is the time difference between adjacent events, and τ is the time constant. After the fit, the live time for a run is estimated as $N \times \tau$ where N is the total number of events in a run.

This method for run 129400 is illustrated in Fig 6.1. Its calculated value is 28730.24 s. The duration for every run is also provided immediately after data taking. The provided duration for this run is 28805 s. The difference in provided duration and

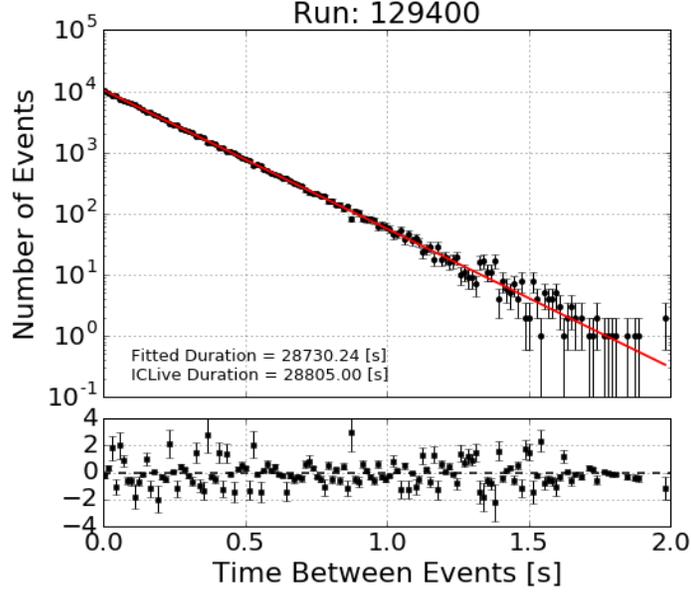


Figure 6.1: An example fit to the histogram of time difference between consecutive events with an exponential decay function. Parameters of the fit is used to calculate duration on run-by-run basis. The difference in duration provided during data taking and calculated by this method is negligible.

estimated duration is negligible. Therefore, this method is used for all runs to get their duration. The total duration (T) from all runs is given by

$$T = \sum_{\text{all runs}} N_{\text{run}} \times \tau_{\text{run}} \quad (6.4)$$

6.3 Effective Area

Simulated showers are re-sampled many times at random locations within a radial distance that depends on the energy of primary particles. Showers with energy 10 TeV to 100 TeV are re-sampled within 200 m from the center of IceTop. Showers with energy 100 TeV up to 1 PeV are re-sampled within 400 m, 1 PeV to 10 PeV are re-sampled within 600 m, and higher than 10 PeV up to 25 PeV are re-sampled within 800 m from the center of IceTop. No events outside these boundary with a given energy can pass IceTop_InFill_STA2 filter. These radial distances from the center of the IceTop geometry are shown in Fig 4.4.

The efficiency of the detector must be known in order to calculate the effective area. Efficiency is the ratio of the number of events that survive all quality cuts to the number of generated events. This ratio is represented by Eq. 6.5.

$$\text{Efficiency}(\epsilon) = \frac{N_{\text{survived}}}{N_{\text{generated}}} \quad (6.5)$$

Effective area is the product of the re-sampled area and the efficiency of the detector and is an energy-dependent quantity. Fig 6.2 shows the effective area and its fit for all-particles with $\cos \theta \geq 0.9$. The left plot of Fig. 6.3 shows the effective area for two different zenith bins and the right plot shows the effective area for proton, iron, and all particles, assuming the H4a composition model. As shown in figures 6.2 and 6.3, the effective area increases and saturates after it reaches a certain energy limit. For example, in Fig. 6.2, the effective area increases from approximately 21000 m² at the lowest energy bin and saturates at approximately 36000 m². The energy region with efficiency higher than 50% is used in this analysis to reduce the dependence of the final energy spectrum on simulations from energy region with lower triggering efficiency.

The effective area is fitted with an energy-dependent sigmoid-like function. It is given by

$$A_{\text{eff}}(E) = \frac{p_0}{1 + e^{-p_1(\log E - p_2)}} \quad (6.6)$$

where $A_{\text{eff}}(E)$ is the effective area and p_0 , p_1 , and p_2 are the fit parameters. Values of these parameters are shown in Appendix A.

6.4 Bayesian Unfolding

One of the challenges that a ground-based detector faces is to connect effect (E , reconstructed energy) seen by the detector to its cause (C , true energy). The goal is to know the true energy distribution (C) from the reconstructed energy distribution (E). An iterative Bayesian unfolding is used and is implemented via a software package called pyUnfolding [124]. This package also calculates and propagates error in each iteration. An example error calculation using this package shows the evolution of systematic uncertainty for 100 iterations (Fig 6.4).

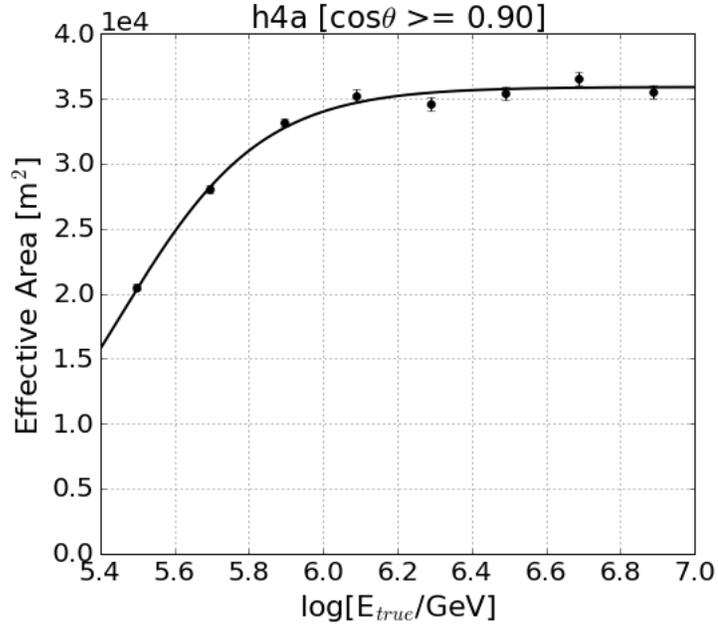


Figure 6.2: Effective area calculated using MC generated with H4a composition model and Sybill 2.1 hadronic interaction model. A sigmoid function is used to fit the effective area.

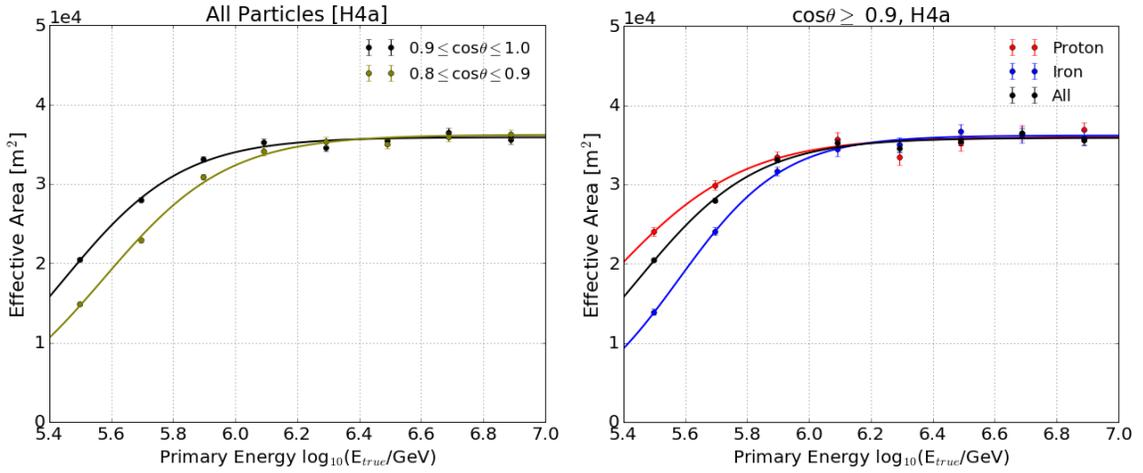


Figure 6.3: Effective area calculated using MC generated with H4a composition model and Sybill 2.1 hadronic interaction model. A sigmoid function is used to fit the effective area. Parameters of all fits are given in Appendix A. Left: Effective area for two zenith bins. Right: Effective area assuming all proton, all iron, and four elements composition.

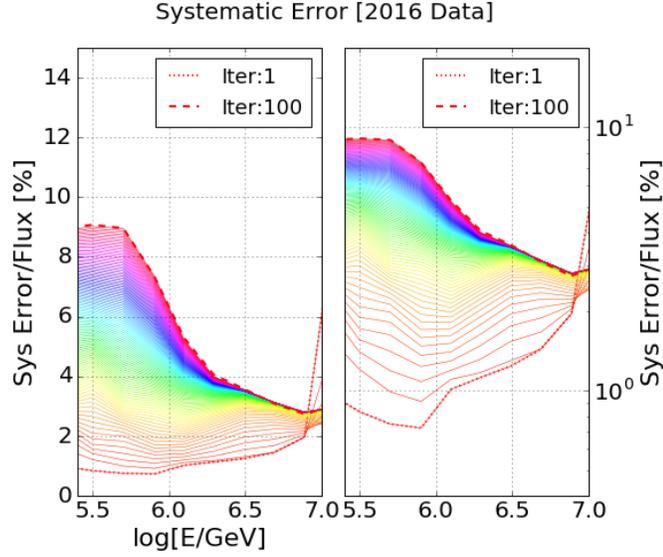


Figure 6.4: An example plot showing an evolution of systematic uncertainty in each iteration for iterative Bayesian unfolding in linear (left) and log (right) y-axis. Iter1 is the bottom line and iter100 is the top line. The mathematical formulation of uncertainty and its propagation in each iteration are given in [123]. The method is implemented in pyUnfolding package.

To unfold the energy spectrum, the response of the detector to an air shower is required. The response is determined from simulations. This information is stored in a response matrix and is the probability of measuring a predicted energy given the true primary energy. Inverting the response matrix to get a probability of measuring true energy given predicted energy would lead to unnatural fluctuations. Therefore, Bayes theorem is used iteratively to get a true distribution from an observed distribution. The response matrix used in this analysis is represented in Fig. 6.5.

The following paragraphs explain D’Agostini’s iterative Bayesian unfolding technique. The Bayes theorem is given by

$$P(C_\mu|E_i) = \frac{P(E_i|C_\mu)P(C_\mu)}{\sum_\nu^{n_C} P(E_i|C_\nu)P(C_\nu)} \quad (6.7)$$

where $P(E_i|C_\mu)$ is the probability that a given cause C_μ results in an effect E_i . $P(E_i|C_\mu)$ is an element of a response matrix. $P(C_\mu)$ is the prior knowledge of the

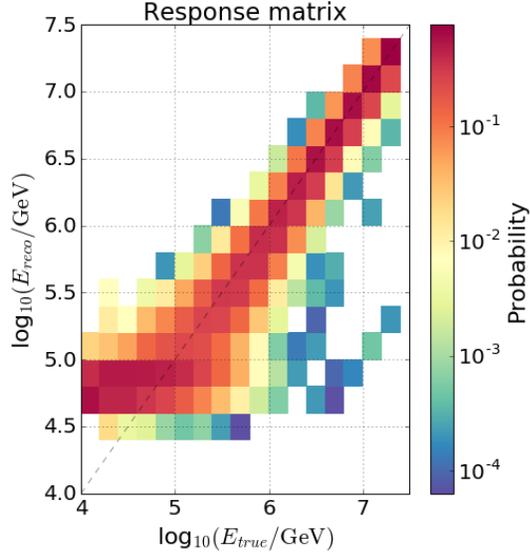


Figure 6.5: Response Matrix calculated from simulation. An element of a response matrix is a fraction of events in true energy bin distributed over the reconstructed energy bin. In Bayes theorem in Eq 6.7, $P(E|C)$ represents a response matrix.

cause distribution. This is the only quantity that changes in the right-hand side of equation 6.7 during each iteration. Each iteration produces a new unfolding matrix $P(C_\mu|E_i)$. $P(C_\mu|E_i)$ represents the probability that an effect E_i is a result of cause C_μ . n_C is the number of cause bins. If the distribution of effect $n(E_i)$ is known then the updated knowledge of the energy distribution is given by

$$\phi(C_\mu) = \frac{1}{\epsilon_\mu} \sum_i^{n_E} P(C_\mu|E_i)n(E_i) \quad (6.8)$$

where $\epsilon_\mu = \sum_j^{n_E} P(E_j|C_\mu)$ and $0 \leq \epsilon_\mu \leq 1$. It must be noted that ϵ_μ in this analysis is equal to 1 and is not the detector efficiency used in Eq. 6.2. $\phi(C_\mu)$ in Eq. 6.8 is used to calculate an updated prior. The updated prior is given by

$$P(C_\mu) = \frac{\phi(C_\mu)}{\sum_\nu \phi(C_\nu)}$$

which is then used as a new prior in equation 6.7 for the next iteration. The condition that $P(C_\mu)$ agrees is $\sum_\mu^{n_C} P(C_\mu) = 1$. The unfolding proceeds until a

desired stopping criterion is satisfied. The commonly used stopping criteria are χ^2 and Kolmogorov-Smirnov of subsequent energy distribution before and after unfolding. In this analysis, Kolmogorov-Smirnov less than 10^{-3} is used as the stopping criterion. At this point, a natural question that arises is what initial prior to use. The choice of initial prior, $P(C_\mu)$, is optional and can be any reasonable distribution, like a uniform distribution or $n(E)/\sum_i^{n_E} n(E_i)$. But, to minimize bias, the conventional choice is Jeffreys' Prior [125] and is given by

$$P_{\text{Jeffreys}}(C_\mu) = \frac{1}{\log_{10}(C_{\text{max}}/C_{\text{min}})C_\mu} \quad (6.9)$$

As mentioned above, the response matrix that has been used for the low energy spectrum analysis is shown in Fig. 6.5. The response matrix is a two-dimensional histogram of events on true and reconstructed energy bin divided by the total number of events in a true energy bin. The response matrix is fixed and does not change throughout the analysis. An unfolding matrix is calculated using the response matrix in the Bayes theorem shown in Eq 6.7. A new unfolding matrix is generated due to the change in the prior in each iteration. In this analysis, the stopping criterion is reached in the twelfth iteration. Unfolding matrices after the first iteration and the last iteration are shown in the top plot of Fig. 6.6.

Each iteration generates a new cause distribution $\phi(C_\mu)$. Using this to calculate the next prior can propagate error, if any, in each iteration which might cause erratic fluctuations on the final distribution. To regularize the process and to avoid passing an unphysical prior in each iteration, the cause distribution ($\phi(C_\mu)$) is fitted with a physical model except on the last iteration.

The energy spectrum follows a power law with a fixed spectral index over a limited energy region. Since the energy region of interest includes the knee, where the flux steepens to a different spectral index, using just one power law might not be the best choice. Therefore, the energy distribution after each iteration as shown in the bottom plot of Fig 6.6 is fitted with a spline fit.

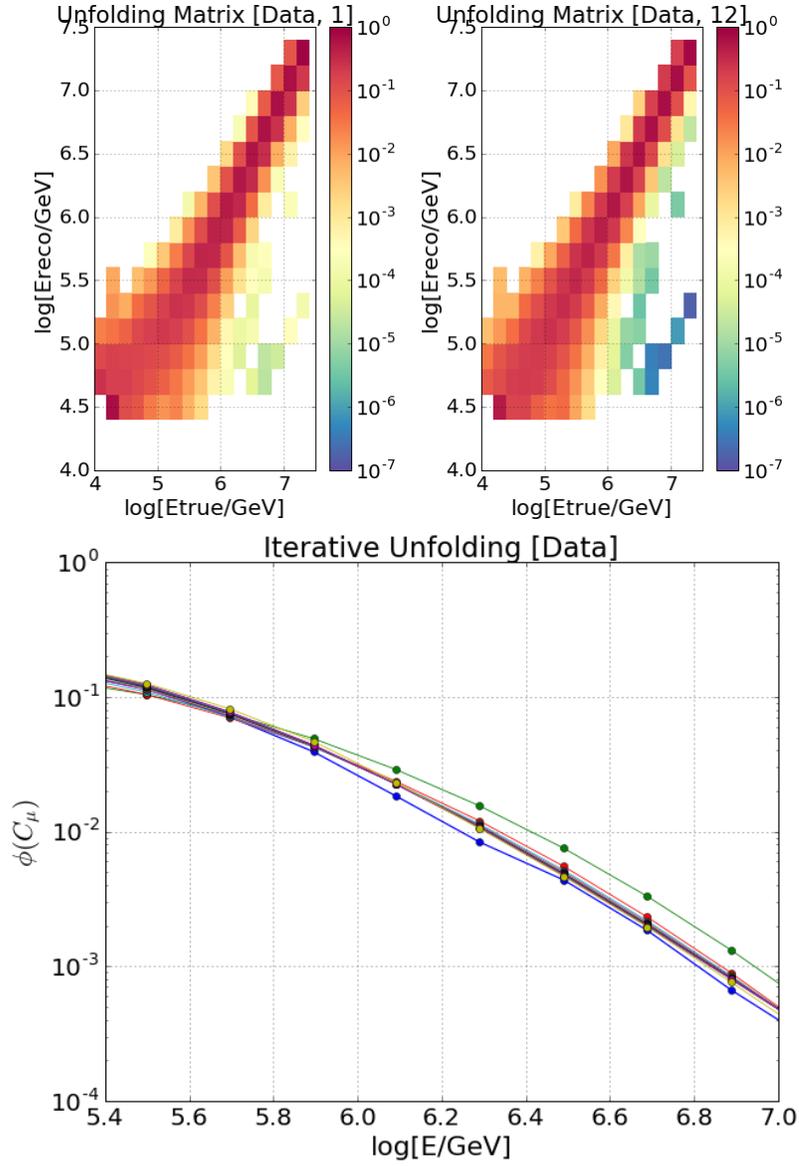


Figure 6.6: Top: Unfolding matrix for the first iteration (left) and the last iteration (right). Bottom: Energy distribution after each iteration of Bayesian iterative unfolding. Iteration was done until termination criteria were met.

6.5 Pressure Correction

The rate of events fluctuates with changes in atmospheric pressure as shown in the left plot of Fig 6.7. If pressure increases, the rate decreases and vice-versa. If the average pressure during which data were taken is not approximately equal to

the pressure of atmospheric profile used in the simulation, then the final flux must be corrected to account for the difference in the atmospheric pressure between data and simulation.

The average pressure for 2016 at the South Pole was 678.27 hPa (data obtained from AMRC). This converts to $1.019 \times 678.27 = 691.16 \text{ g/cm}^2$. In CORSIKA, the density variation of the atmosphere with altitude is modeled in 5 layers. At the altitude h (cm) from sea level, the overburden $T(h)$ of the atmosphere is given by the form

$$T(h) = a + b \exp\left(-\frac{h}{c}\right) \quad (6.10)$$

where h is the altitude of detector in unit of cm, $T(h)$ is the overburden at altitude h in units of g/cm^2 . For the energy region of interest in this analysis, the average April atmosphere was used in CORSIKA. Parameters used for average April atmosphere are $a = -69.7259$, $b = 1111.7$, and $c = 766099$. For IceTop at an altitude of 2835 m from sea level, using these parameters, $T(h)$ in the simulation was $\sim 1\%$ larger than the average pressure for the period of data taking ($698.12/691.1 \text{ g/cm}^2$).

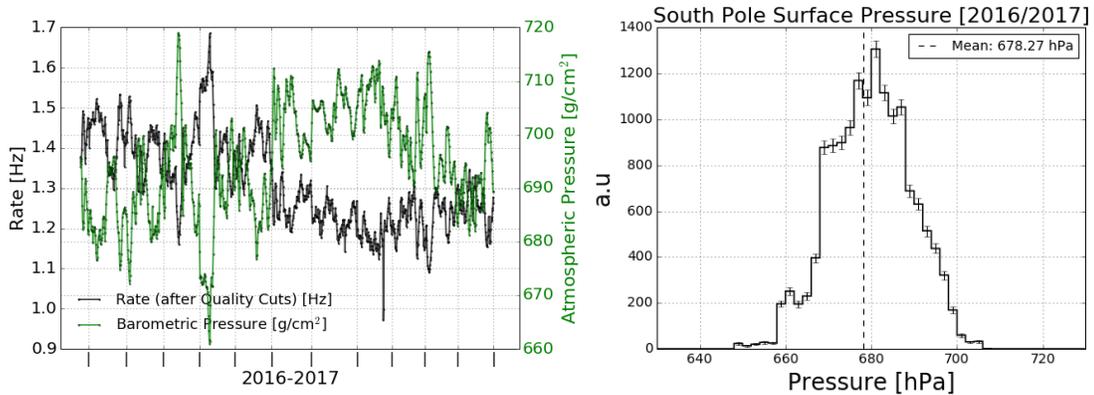


Figure 6.7: Left: The rate of two-station events on run-by-run basis compared with the average South Pole atmospheric pressure during data-taking period. Right: Pressure histogram for 2016 South Pole atmosphere with a mean of $\sim 691 \text{ g/cm}^2$.

To get data with an equivalent to 698 g/cm^2 , 38 runs from 13 days of data (Jan 10-20 2017 and Mar 03-06 2017) are used. The flux for these 38 runs is calculated

and compared with the flux using 2016 data. The flux decreases with an increase in pressure and this decrease must be corrected. Fig. 6.8 shows the correction factors and these are tabulated in Table 6.1. Errors on the correction factors due to pressure difference are used as the systematic uncertainty on flux due to atmospheric pressure.

Table 6.1: Correction factor on the final flux due to difference in atmospheric pressure between simulation and 2016 data.

$\log[E/\text{GeV}]$	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Low [%]	4.93	5.13	5.54	5.86	5.56	5.54	6.18	5.43
High [%]	4.93	5.13	5.54	5.86	5.56	5.54	6.18	5.43

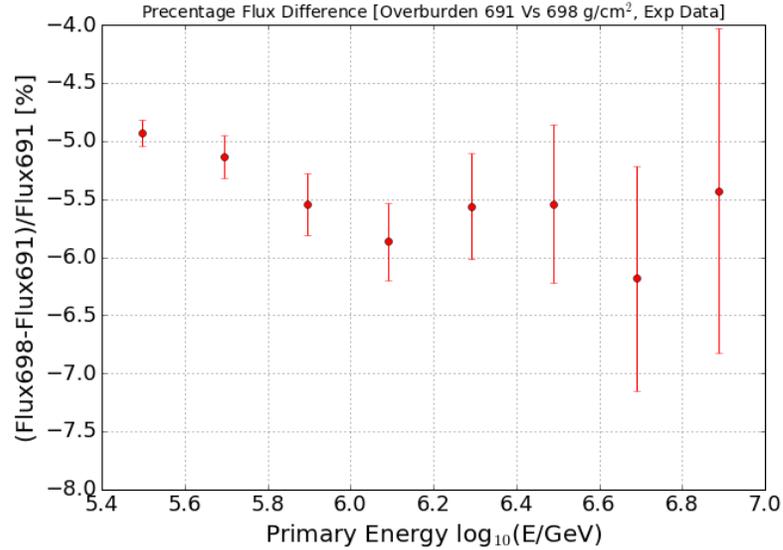


Figure 6.8: Percentage deviation of cosmic rays flux when atmospheric pressure is $\sim 698 \text{ g/cm}^2$ from the flux when pressure is $\sim 691 \text{ g/cm}^2$. This deviation is used as the correction factor to correct the final flux. The error on the correction factor is used as the systematic uncertainty due to Pressure difference between average pressure of 2016 South Pole atmosphere and pressure due to atmosphere profile used in simulation.

6.6 Systematic Uncertainties

Systematic uncertainties are all uncertainties other than statistical uncertainties. When calculating different systematics, all conditions except the systematic under investigation are held constant. There are several sources of systematic uncertainties in this analysis. The sources are listed below of which there are four primary sources and some negligible sources. The primary sources are:

- Uncertainty due to the composition.
- Uncertainty due to the unfolding method.
- Uncertainty due to the effective area.
- Uncertainty due to the atmospheric pressure.

and the negligible sources are:

- Uncertainty due to VEM Calibration.
- Uncertainty due to snow.
- Uncertainty due to duration.

6.6.1 Uncertainty due to Composition

The Gaisser H4a model, shown in Fig 4.5, is used as an assumption for the composition. The Gaisser model is more realistic than simply comparing the extremes: pure proton and pure iron. Three other models are used to calculate systematic uncertainty due to the composition. They are GST, GSF, and a modified version of Polygonato. Additionally, the fractional difference between fluxes calculated for two extreme zenith bins, as used in the 3-year energy spectrum analysis, is used to calculate composition systematics.

To reconstruct the energy of events and eventually to calculate flux, the random forest model assuming H4a is trained and the energy for each event is predicted. Then a response matrix is generated using the baseline H4a composition model to calculate flux

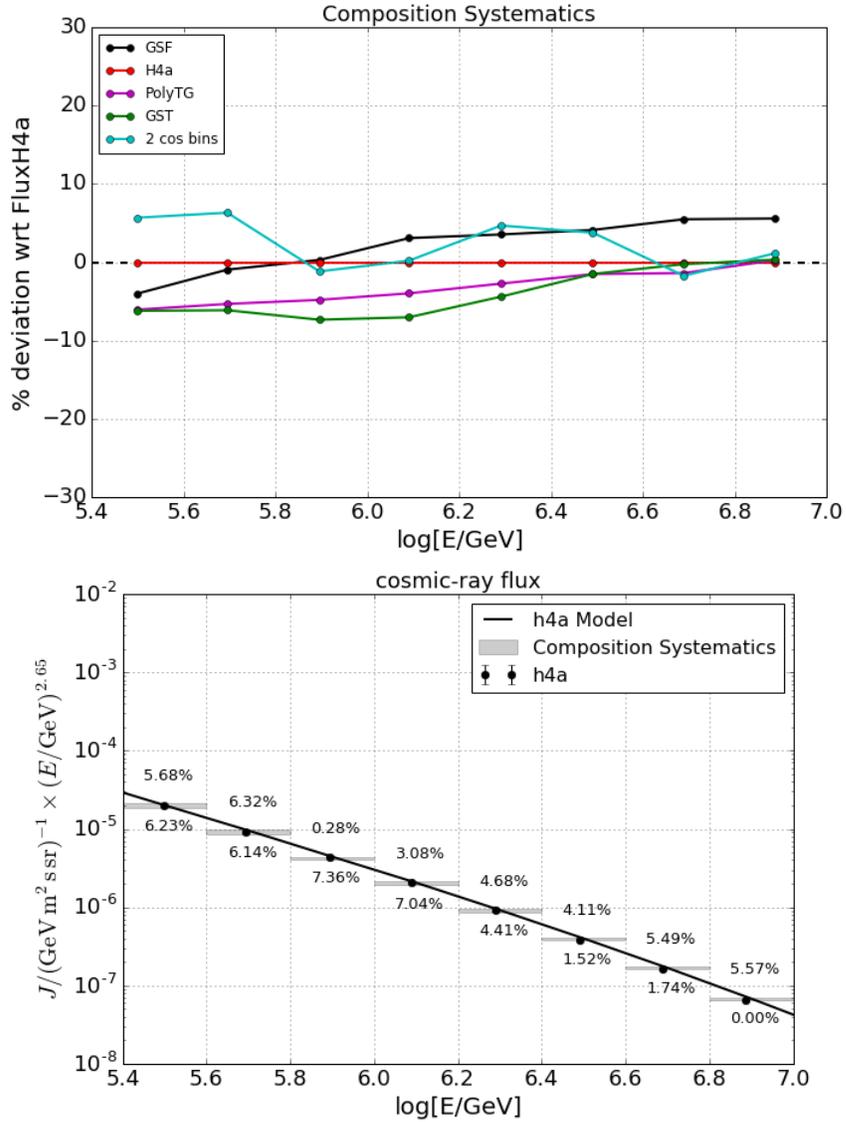


Figure 6.9: Top: Percentage spread of cosmic ray flux assuming different composition models. Maximum spread is used as the systematic uncertainty due to composition. Bottom: Systematic uncertainty due to composition relative to the H4a composition model.

for all composition models. Since all these models are viable options for composition, the flux for each model is calculated and the percentage deviation of flux from the model itself for each energy bin is measured. The deviation of output flux from individual models in percentage terms is shown in the top plot of Fig 6.9. The maximum spread

of deviation is used as the uncertainty due to composition and is shown in the bottom plot of Fig 6.9.

6.6.2 Uncertainty from Unfolding

The pyUnfolding package calculates the systematic uncertainty due to unfolding at the end of each iteration. Evolution of systematic uncertainty after each iteration is saved and is shown in Fig 6.10. These are the fraction of systematic uncertainty to unfolded flux for each iteration. The analysis is performed using the H4a composition model and the Jeffreys prior. In this study, the number of iterations is twelve for both data and simulation before reaching the termination criterion. The systematic uncertainty for the twelfth iteration is used as the systematic uncertainty due to the unfolding procedure. Fig 6.10 shows the systematic uncertainty for each iteration and Table 6.2 shows the values for the twelfth iteration.

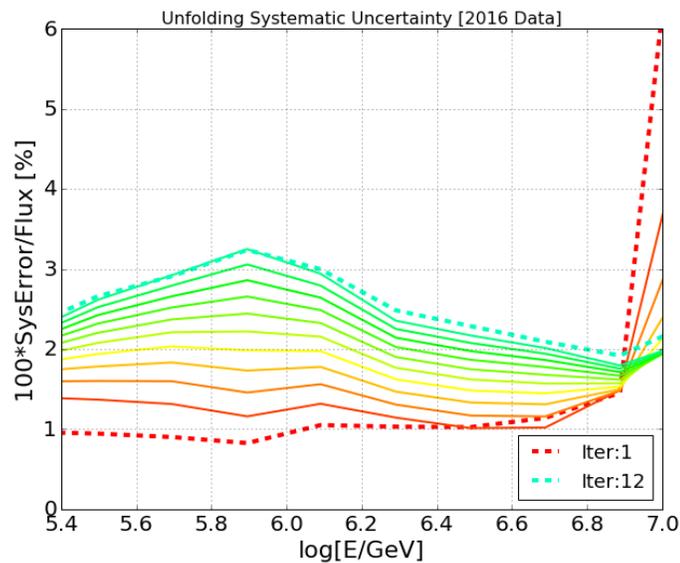


Figure 6.10: Systematic uncertainty for each of twelve iterations in iterative Bayesian unfolding. The twelfth systematic uncertainty is used as the systematic uncertainty from unfolding procedure. The mathematical formulation of uncertainty and its propagation in each iteration is shown in [123]. The method is implemented in pyUnfolding package.

Table 6.2: Systematic uncertainty due to iterative Bayesian unfolding.

$\log[E/\text{GeV}]$	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Low [%]	2.66	2.91	3.24	2.99	2.48	2.29	2.09	1.92
High [%]	2.66	2.91	3.24	2.99	2.48	2.29	2.09	1.92

6.6.3 Uncertainty from Effective Area

The effective area is fitted with an energy-dependent function given in Eq. 6.6. The parameters of the fit contain errors and the errors have to be accounted for while calculating flux. A band around the effective area fit is shown in Fig. 6.11 after accounting for all errors on the parameters. Taking the upper and lower boundary of the band, the flux is calculated and the difference in the flux is used as the systematic uncertainty due to effective area. Values of systematic uncertainty as a percentage of the fit are shown in Table 6.3.

Table 6.3: Systematic uncertainty due to effective area. Uncertainties for each energy bin are calculated with respect to the fit.

$\log[E/\text{GeV}]$	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Low [%]	2.56	3.57	2.71	1.99	1.45	1.18	1.06	1.01
High [%]	1.61	2.16	2.02	1.46	1.20	1.06	0.99	0.97

6.6.4 Uncertainty due to Pressure

The correction factor to account for the atmospheric pressure difference between data and simulation is shown in Fig 6.8. The uncertainty on the correction factor is used as the systematic uncertainty due to pressure. The values of systematic uncertainty are shown in Table 6.4.

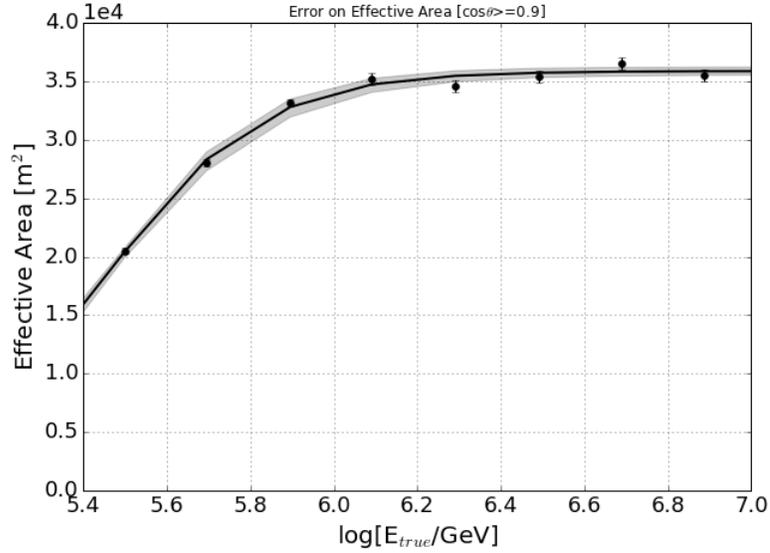


Figure 6.11: Effective area calculated using MC generated with Sibyll2.1 hadronic interaction model and using H4a as composition model. Solid black line is the fit with Equation 6.6 and a gray band is an error of the fit. Energy region with efficiency higher than 50% is used in this analysis.

Table 6.4: Systematic uncertainty due to difference in atmospheric pressure.

$\log[E/\text{GeV}]$	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Low [%]	0.113	0.181	0.266	0.335	0.457	0.679	0.969	1.395
High [%]	0.113	0.181	0.266	0.335	0.457	0.679	0.969	1.395

6.6.5 Uncertainty in VEM Calibration

Charge on tanks from an air shower is extensively used in this analysis. The charge on a tank depends on the amount of deposited energy by secondary particles of a shower. A systematic uncertainty can arise if the calibration on the unit of charge (VEM) is not consistent between data and simulation.

The simulation uses October 2016 VEMCal values. If the mean of calibration for 2016 is not equal to the values used in the simulation, a systematic effect arises. To check the agreement of VEMCal value between data and simulation, a histogram of difference between the mean value for 2016 and value used in the simulation is

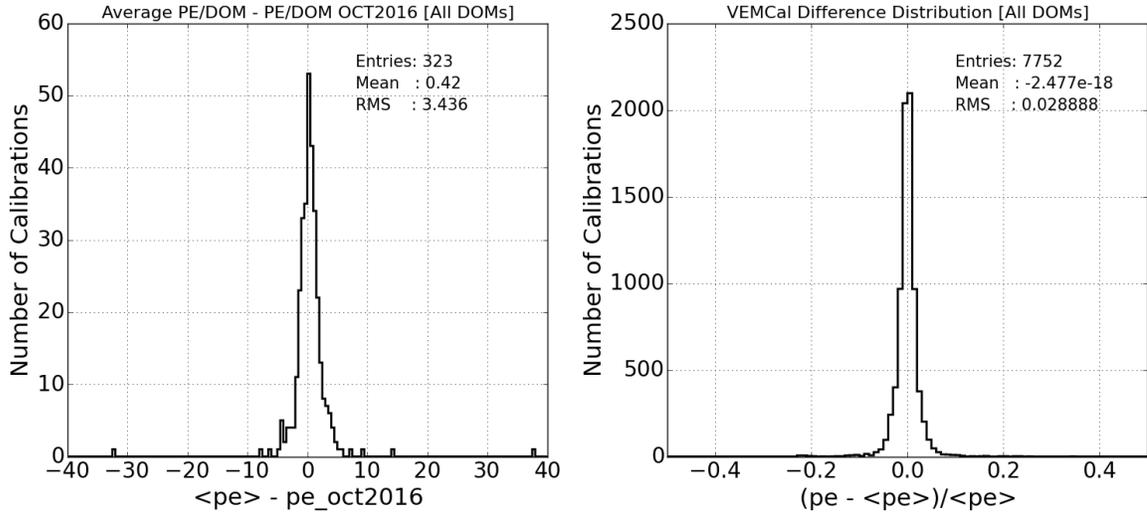


Figure 6.12: Left: Histogram of difference between mean value of VEMCal for 2016 in the unit of pe and value used in simulation. The difference is calculated on DOM-by-DOM basis. Right: Distribution of deviation of 1 VEM in pe for 2016 from its mean on DOM-by-DOM basis.

calculated. As shown in the left plot of Fig 6.12, the mean of the shift is only 0.42 pe. The distribution of deviation of VEM value from its mean is shown on the right plot of Fig 6.12. This distribution has an RMS of 0.028, which is almost 3%.

One VEM is approximately equal to 120 pe. Since the mean value of VEMCal and values used in simulation varies by only 0.42 pe, we considered this value as a negligible difference. Hence, the systematic uncertainty due to VEM calibration is ignored.

6.6.6 Uncertainty due to Snow

Different snow heights for data and simulation affect the low energy spectrum analysis. For simulation, October 2016 snow heights have been used. In Fig 6.13, October 2016 snow heights are superimposed on the IceTop geometry. In Fig 4.2, the comparison between average 2016 snow heights and October 2016 snow heights is shown for each station. The average amount of snow per tank is 1.747 m for 2016 data. The average snow height per tank is 1.764 m for simulation. On average, the amount

of snow difference per tank between data and simulation is 0.017 m.

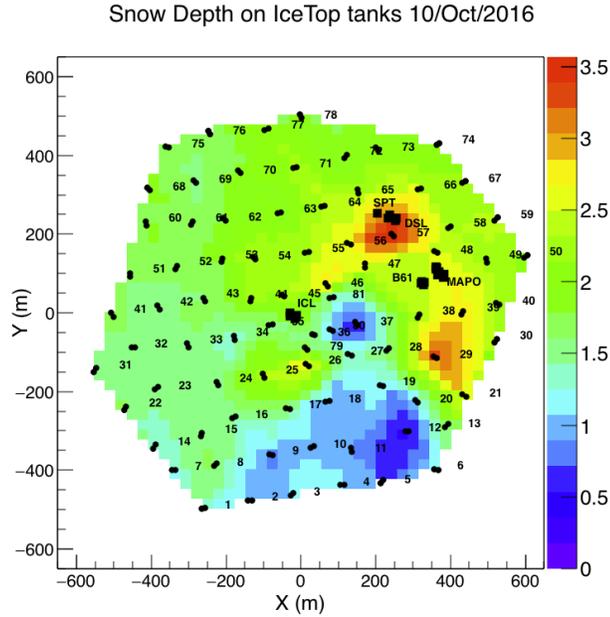


Figure 6.13: Snow height on top of IceTop geometry. Snow height on top of each station is shown in Fig 4.2. To compare snow heights between data and simulation, an average value of 2016 is used for data and October 2016 snow height is used for simulation.

Since the average height of snow between data and simulation is comparable and the dispersion is negligible, the systematic uncertainty due to snow is ignored in this analysis.

6.6.7 Uncertainty due to Duration

Duration is calculated by fitting an exponential decay function on the histogram of time intervals between two consecutive events. The fit is performed on a run-by-run basis. The fit parameters have errors and cause the upper and lower limit on the nominal duration for 2016 data. The nominal duration for 2016 runs is 28548809.85 s. The lower limit on the duration is 28487813.25 s whereas the upper limit is 28609806.46 s. The difference in the flux using these upper and lower limits

compared to the nominal flux is 0.21%. Hence, the systematic uncertainty due to duration is negligible and is ignored in this analysis.

6.6.8 Total Systematic Uncertainty

In this analysis, the systematic uncertainty due to composition, unfolding method, effective area, and pressure play a vital role while other systematics have negligible results. After calculating all contributions to the systematic uncertainty, we found that it is dominated by composition. The total systematic uncertainty is the sum of these individual systematic uncertainties in quadrature. Individual systematic uncertainties along with the total are shown in Fig. 6.14. The total systematic uncertainty for each energy bin is also tabulated and is shown in Table 6.5.

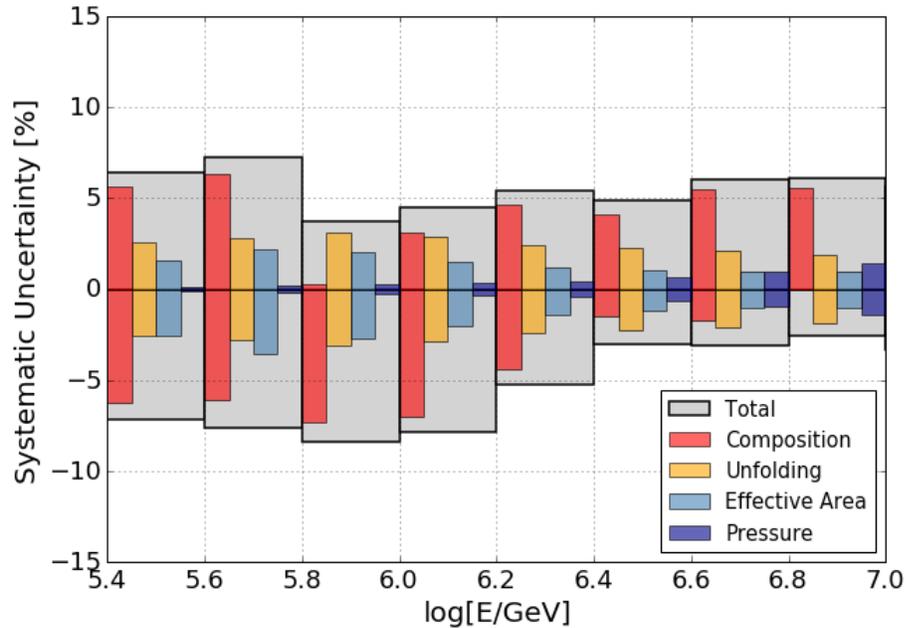


Figure 6.14: The plot shows the individual systematic uncertainties for each energy bins. Total systematic uncertainty is the sum of individual uncertainty in quadrature.

Table 6.5: Total systematic uncertainty after adding individual systematic uncertainty in quadrature.

$\log[E/\text{GeV}]$	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Low [%]	7.29	7.75	8.57	7.99	5.36	3.17	3.18	2.67
High [%]	6.54	7.36	4.01	4.68	5.53	4.94	6.09	6.18

6.7 Statistical Uncertainties

As defined in [126], “the variance, or the standard deviation, of the estimator is a measure of how widely the estimates would be distributed if the experiment were to be repeated many times with the same number of observations per experiment.” It is common in the literature to report standard deviation σ of an underlying distribution as the statistical uncertainty of measurement.

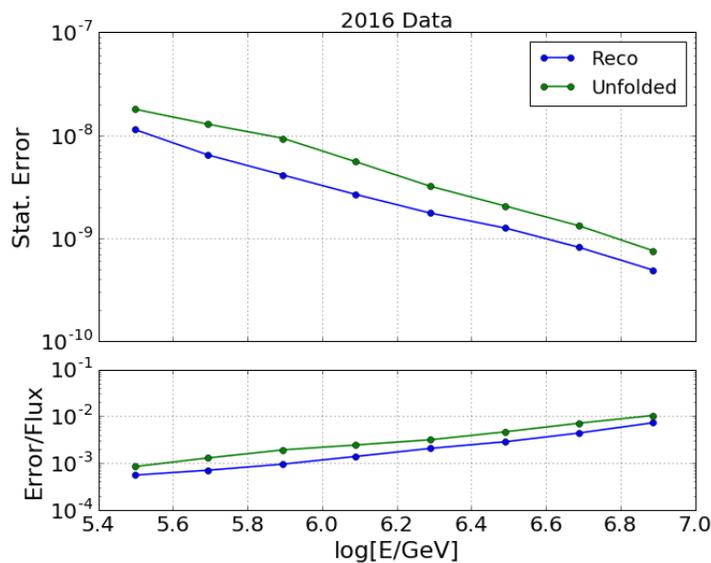


Figure 6.15: Statistical error calculated using initial energy distribution assuming it follows a Poisson distribution and an output after Bayesian iterative unfolding. The statistical error calculated using the pyUnfolding package is used as the final statistical uncertainty which is higher but still small compared to the flux.

In iterative Bayesian unfolding, the statistical error changes in each iteration as the number of events in each bin changes. The pyUnfolding package keeps track of

energy distributions at each iteration and calculates final statistical uncertainty. The statistical error assuming a Poisson distribution of events in each reconstructed energy bin and the error calculated using the pyUnfolding package are shown in Fig 6.15. The bottom plot of Fig 6.15 shows the ratio of statistical error to its flux. It can be seen that the statistical error is more than two orders of magnitude smaller than the flux and more than an order smaller than the total systematic uncertainty.

6.8 Cosmic Ray Flux

Here we discuss the main result of this analysis, which lowers the threshold for the IceTop energy spectrum by an order of magnitude. The data were obtained with a new trigger designed to include events with only two stations hit. Data from May 2016 to April 2017 are used.

As described, random forest regression is used for reconstruction of core position, zenith, and energy. Only events with zenith less than 26° are used as they are better reconstructed. The cosmic ray flux in the energy range 250 TeV to 10 PeV is derived from events with zenith angle range $0^\circ < \theta < 26^\circ$. The reconstructed events are placed in energy bins and iterative Bayesian unfolding is performed using the pyUnfolding package. The unfolded energy distribution is divided by exposure to calculate the flux using Eq. 6.3. The all-particle cosmic ray energy spectrum measured using the IceTop detector with 81 stations configuration is shown in Fig 6.16. The top plot of Fig 6.16 shows the flux, $J(E)$, of cosmic rays in unit of $\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}$ and the bottom plot shows the energy spectrum scaled by $E^{1.65}$, i.e $E^{1.65} \times J(E)$.

The statistical and systematic uncertainty of the flux are calculated by using the pyUnfolding package. The statistical uncertainty is small due to the large volume of data. The systematic uncertainty from the composition assumption is the largest, whereas, the systematic uncertainty from unfolding method, effective area, and pressure are relatively small. The total systematic uncertainty is calculated by adding individual uncertainties in quadrature. The total systematic uncertainty is larger than the total statistical uncertainty.

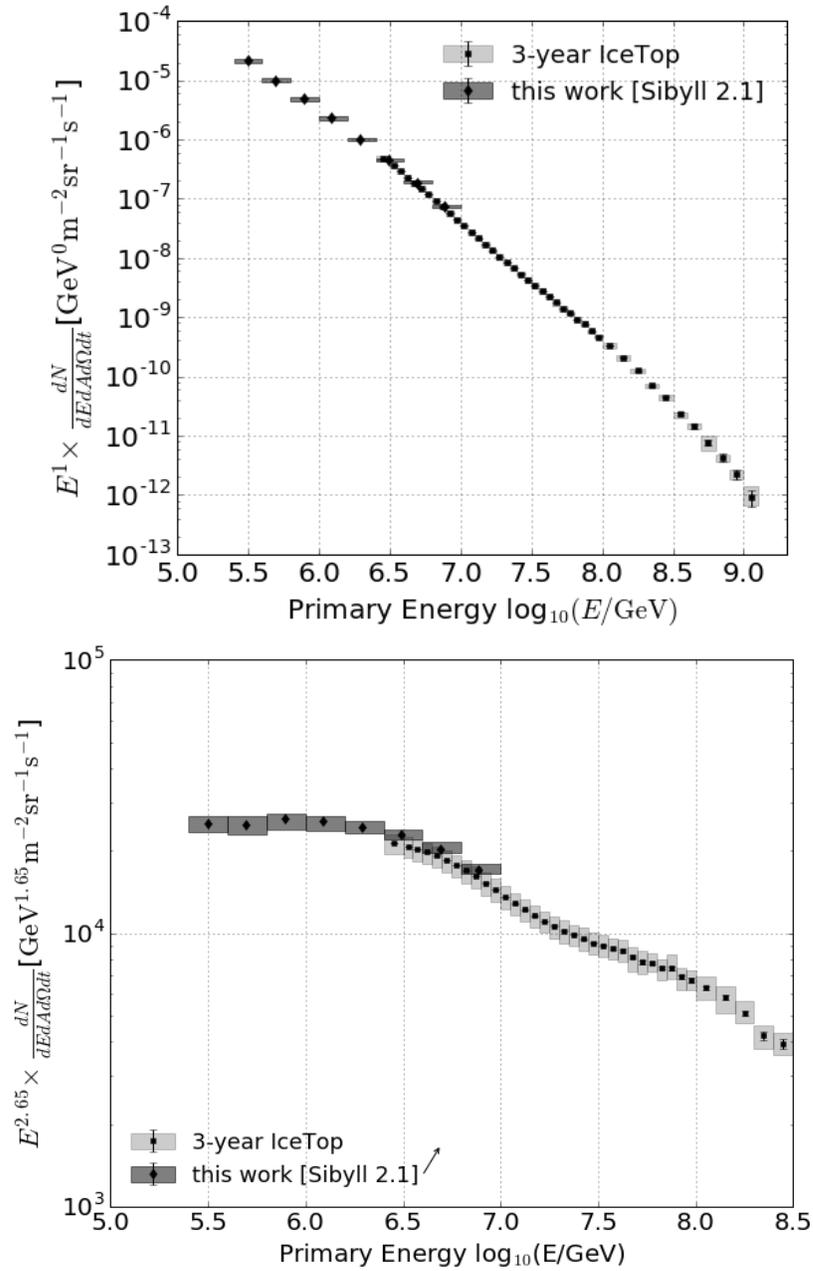


Figure 6.16: Top: Cosmic Ray Flux using IceTop 2016 Data. Bottom: Cosmic Ray Flux using IceTop 2016 Data scaled by $E^{1.65}$.

Many ground-based cosmic ray detectors measure the cosmic ray flux around this energy region. Several measurements with their statistical uncertainties are compared with the low energy IceTop flux in Fig 6.17. The range of results reflects

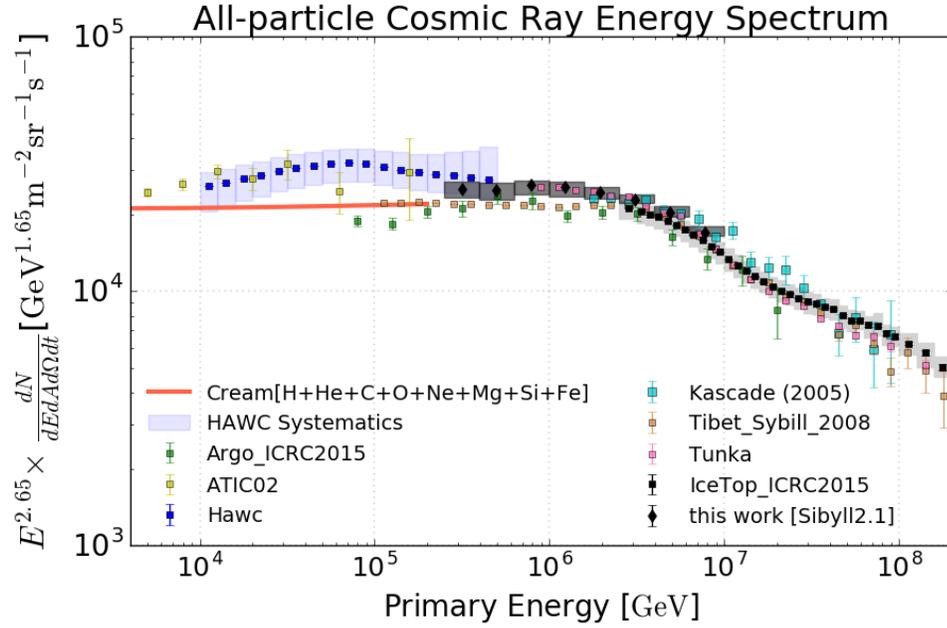


Figure 6.17: Cosmic Ray Flux using IceTop 2016 Data compared with other experiments. Special thanks to Stephane Coutu for providing with the fit parameters for CREAM energy spectrum.

systematic uncertainties in the measurements. The cosmic ray flux follows a power law and a slight difference in energy scale can cause a large difference in the flux. Additionally, in Fig 6.17 the spectrum is scaled with $E^{1.65}$ which magnifies even a small flux difference.

The IceTop low energy spectrum extension overlaps with the results from *High Altitude Water Cherenkov* (HAWC) [80] detector at lower energy region and with Cascade [74] and Tunka [76] detectors at higher energies. The IceTop low energy spectrum is also compared with direct measurements from ATIC-02 [69] and CREAM [68].

6.9 Tests

In order to test the robustness of our measurement, fluxes are calculated assuming several bin centers and initial priors for Bayesian unfolding. Given that enough data is available, iterative Bayesian unfolding should result in a flux that

agrees within systematics independent of bin size. Results for different bin centers are compared with each other and results for different priors are compared with each other. Also, results from two different energy bin sizes are compared.

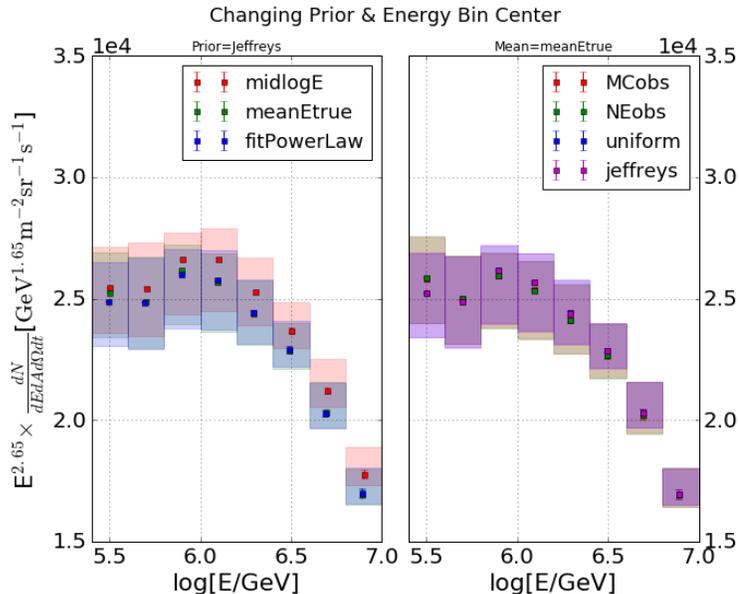


Figure 6.18: Left: Effect of different choice of bin center on the cosmic ray flux. Right: Effect of different priors on the cosmic ray flux

6.9.1 Different Mean Energy and Initial Prior

The energy bin center is used to scale the final cosmic ray flux by $E^{1.65}$. Three different techniques have been studied to calculate the energy bin centers. The easiest method to calculate the mean energy bin is to use the bin center in $\log E$ bins. The second method is to fit the cosmic ray flux to find the spectral index. Using the spectral index, the bin center is defined as the energy in a bin that divides the total number of events in that specific bin in half. The final method to calculate the bin center is to use the mean of the true energy of events for each energy bin. Fluxes from these methods are within the total systematic uncertainty. The bin centers calculated by the mean of true energy of events and by fitting cosmic ray flux are close to each other, dividing the number of events in that energy bin to half. Hence, the mean of true energy of

events for each energy bin is used as the bin center. The flux calculated from these three methods are not significantly different as can be seen in the left plot of Fig 6.18.

An initial prior is used in the iterative Bayesian unfolding. Four different priors are studied in this analysis: uniform prior, Jeffreys' prior, normalized reconstructed energy spectrum from data and from MC. Jeffreys' prior is used in this analysis as it is commonly used as an unbiased prior. Comparison of the final flux assuming these four priors is shown in the right plot of Fig 6.18. As can be seen, the fluxes are within systematic uncertainty.

6.9.2 Different Bin Size

One of the benefits of using iterative Bayesian unfolding is that it accounts for bin migration. Previous studies [101] account for only adjacent bin migrations and the minimum bin size used in those studies is close to the energy resolution. Since iterative Bayesian unfolding accounts for all bin migrations, this study is not limited by the size of energy bins. If the amount of data is large enough, then a smaller energy bin size can be used. Fig 6.19 compares the cosmic ray flux for a bin size of $\Delta \log_{10}[E/\text{GeV}] = 0.1$ with $\Delta \log_{10}[E/\text{GeV}] = 0.2$. There is not much difference between fluxes and the small difference that is present is within the systematic uncertainties.

The bin size used in this analysis, $\Delta \log_{10}[E/\text{GeV}] = 0.2$, is close to the energy resolution. Flux calculated using $\Delta \log_{10}[E/\text{GeV}] = 0.1$ is shown as proof that iterative Bayesian unfolding accounts for all energy bin migrations and any bin size can be used if there is enough data.

6.10 QGSJetII-04 as Hadronic Interaction Model

In order to estimate the effect of the interaction model, Sibyll 2.1, that we have used on all-particle cosmic ray flux, a parallel analysis using simulation with QGSJetII-04 as hadronic interaction model has been done. Due to technical limitations, a smaller set of simulations using QGSJetII-04 is produced. QGSJetII-04 simulations are 10% of Sibyll 2.1 simulations.

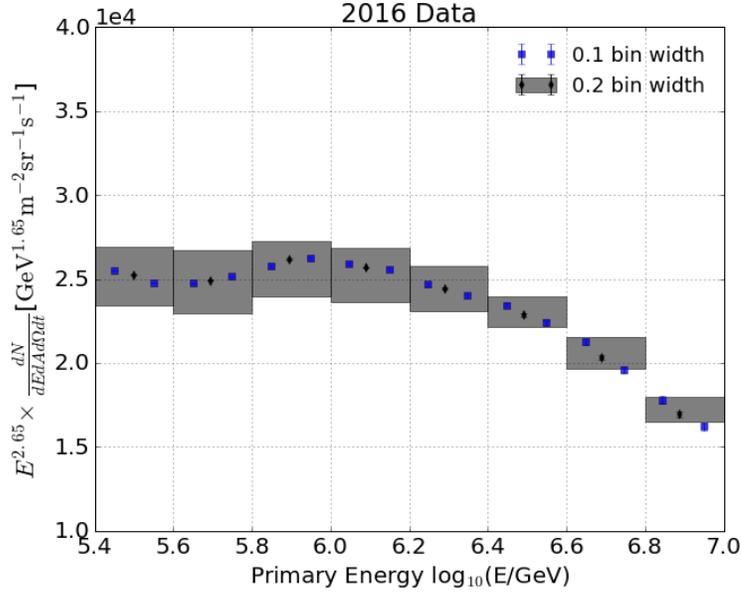


Figure 6.19: All-particle cosmic ray flux for two energy bins to check the effect of different bin size. Iterative Bayesian unfolding accounts for all energy bin migration proving the choice of bin size is arbitrary if there is enough data.

For comparison purposes, the same reconstruction method used for events with Sibyll is repeated for events with QGSJet. As a reminder, random forest regression is used to reconstruct core position, zenith, and energy. For the composition model, H4a is used. A model generated by training on simulation with QGSJet is used to predict energy for experimental data. The energy spectrum predicted this way is unfolded using iterative Bayesian unfolding. Once the energy spectrum is unfolded, this spectrum is used to calculate flux using Eq.6.3. Once the flux with QGSJet is calculated, it is compared with Sibyll2.1 flux. The same steps used to calculate systematic uncertainties for Sibyll 2.1 are used for QGSJet.

Fig. 6.20 shows the results for all-particle cosmic ray flux using QGSJetII-04 as the hadronic interaction model. The top plot shows the flux along with the 3-year IceTop spectrum. It can be seen that the flux with QGSJetII-04 as hadronic interaction model assumption is lower than that of the 3-year spectrum at the overlap region. On the bottom plot, the results of the two-station, low-energy analysis are compared for

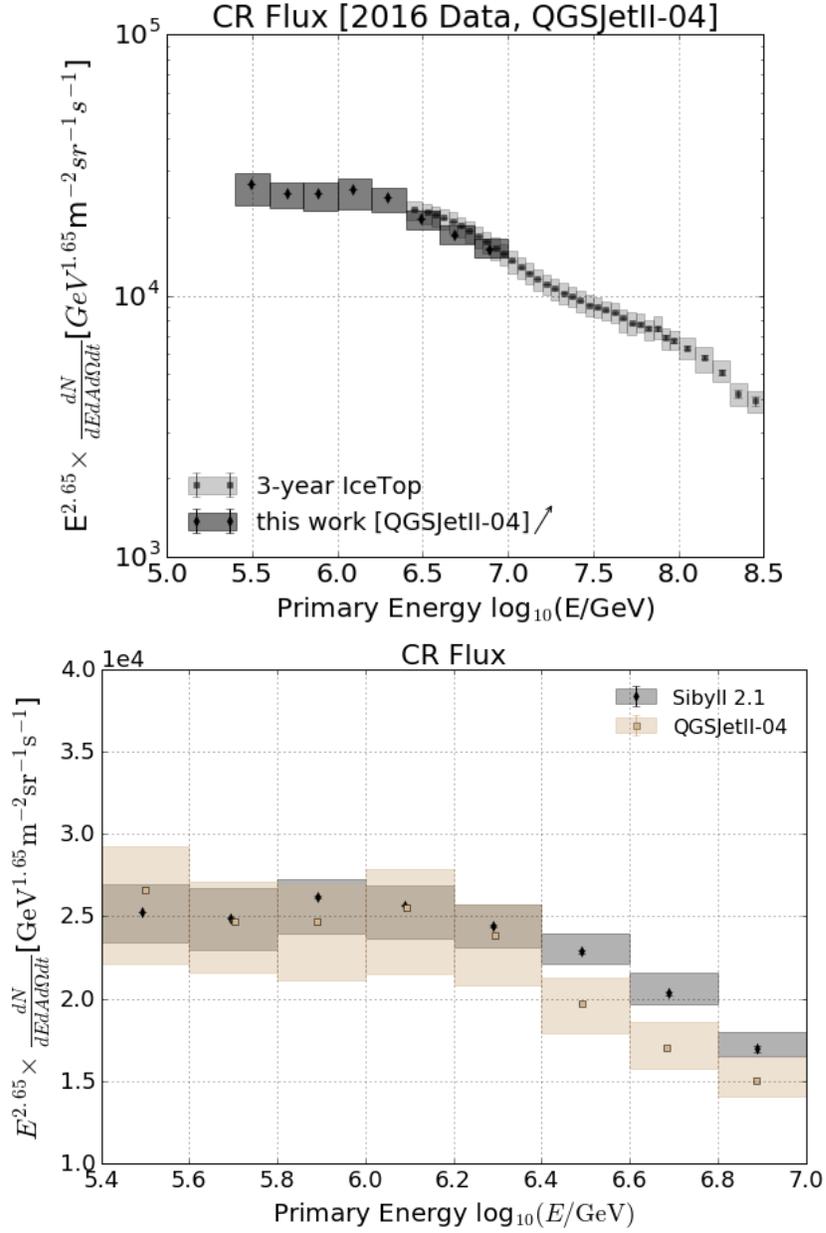


Figure 6.20: Top: Scaled cosmic ray flux assuming QGSJetII-04 hadronic interaction model and compared with 3-year IceTop energy spectrum. The gray band around the flux is the total systematic uncertainty. Bottom: Comparison of the final flux assuming Sibyll2.1 and QGSJetII-04.

the two hadronic interaction models. The flux assuming QGSJet overlaps below the PeV region and is lower than Sibyll 2.1 at higher energy.

Difference in energy spectra due to different hadronic interaction models is not treated as systematic uncertainty in this analysis. Systematics due to hadronic interaction model in this case is mixed with other systematics like systematic uncertainty due to unfolding. Therefore, the results are shown as two separate analyses in the bottom plot of Fig. 6.20. The study of effects of different hadronic interaction model on cosmic ray energy spectrum is found in [81, 127].

Chapter 7

CONSTANT INTENSITY CUT METHOD

The Constant Intensity Cut (CIC) method is a tool that can be used to calculate the cosmic ray energy spectrum. It has also been used to estimate the proton-air interaction cross section [128]. This chapter discusses how the CIC method can be implemented on IceTop data to calculate the cosmic ray energy spectrum. The method is also used to estimate mean shower maximum indirectly by fitting the attenuation curve with a Gaisser-Hillas type function.

7.1 Motivation

Signals on the ground caused by particles with the same primary energy decrease with increasing zenith angle. Particles have to travel through more atmosphere as the zenith angle increases resulting in a loss of many secondary particles from the extensive air shower. This effect is called atmospheric attenuation. It is important to correct for the attenuation on showers to compare detected signals coming from various directions. The correction improves the reconstruction of the primary energy of events from signals detected on the ground. One way to account for the attenuation is to use the CIC method.

The basic assumption of the CIC method is that the cosmic ray flux is isotropic. Isotropy implies that the arrival frequency of cosmic rays depends only on the primary energy and not on the arrival direction, establishing a direct relation between primary energy and the number of events. The intensity of particles with a given primary energy must be constant in any direction.

Previous IceCube spectrum measurements have used events with zenith angle less than 37° [98, 99, 100, 101]. The CIC method enables the use of events with

higher zenith angle that leads to an increase in the statistics. The increase in statistics reduces the statistical uncertainty of the cosmic ray flux at higher energy bins. During the process, the mean shower maximum as a function of energy can also be derived.

7.2 Shower Reconstruction

The basic technique in the CIC method is to get a spectrum of energy proxy, S_{125} , for various angular bins and to cut on a constant rate of events. S_{125} is the signal at 125 m from the shower axis. The decrease in signal size at higher zenith angle for constant number of events shows the attenuation of the shower as a function of slant depth. The following paragraphs discuss the reconstruction technique used to get S_{125} and shower zenith angle.

IceTop gathers information on the charge and time of a hit in a tank. Since the position of the hit tank is known, charge, time of hit, and position of hit tanks are used for reconstruction of S_{125} and direction using Laputop. Shower direction, core position, and shower size are reconstructed by fitting the charges in each tank with a Lateral Distribution Function (LDF) and the signal times with a function describing the geometrical shape of the shower front. The LDF is given by

$$S(R) = S_{\text{ref}} \left(\frac{R}{R_{\text{ref}}} \right)^{-\beta - \kappa \log_{10} \left(\frac{R}{R_{\text{ref}}} \right)} \quad (7.1)$$

where S_{ref} is the shower size at the reference distance R_{ref} to the shower axis, and β is the slope of the logarithmic LDF at R_{ref} . R_{ref} is 125 m in the standard reconstruction. The shower front is described by a plane shower front with a Gaussian parabolic nose. The time of hit on each tank is described by

$$t(x) = t_0 + \frac{1}{c} (\vec{x} - \vec{x}_c) \cdot \hat{n} + \Delta t(R) \quad (7.2)$$

where \vec{x} is the position of the tank, \vec{x}_c is the position of shower core, and \hat{n} is the unit vector in the direction of the air shower. The Gaussian parabolic nose of a shower front is represented by $\Delta t(R)$ and is given by

$$\Delta t(R) = aR^2 + b \left(1 - e^{-\frac{R^2}{2\sigma^2}} \right) \quad (7.3)$$

where $a=4.823\times 10^{-4}$ ns/m², $b = 19.41$ ns, and $\sigma=83.5$ m [87].

Equation 7.1 and 7.2 describe the expectations for the lateral charge distribution and time of hit. A maximum likelihood method is used to estimate S_{125} , \vec{x}_c , β , t_0 , and \hat{n} . In addition, the information of tanks that do not trigger is also used in the maximum likelihood method. The shower size, S_{125} , is defined as the fitted value of the LDF at a reference distance of 125 m away from the shower axis.

Since IceTop is located at the South Pole, an average of 20 cm snow is accumulated on top of each tank every year. The snow attenuates the shower before it reaches the tank. To correct for this attenuation, the expected signal in the likelihood fitting procedure is reduced for each tank according to

$$S_{\text{with snow}} = S_{\text{no snow}} e^{-\frac{d \sec \theta}{\lambda}} \quad (7.4)$$

where d is the height of snow on top of the tank, θ is the reconstructed zenith angle of the shower and $\lambda=2.25$ m is the effective attenuation length of the electromagnetic component of the shower in the snow. The S_{125} used for the constant intensity cut method is the snow corrected signal.

The relation between S_{125} and primary energy used in IceTop-alone 3-year spectrum analysis is given by

$$\log_{10} E = p_0 + p_1 \log_{10} S_{125} \quad (7.5)$$

The S_{125} to energy conversion relation for H4a composition in the zenith bin of $\cos \theta$ 0.85-0.9 is given by Eq 7.5 where, $p_0 = 6.109777$, and $p_1 = 0.914971$. The result of all-particle cosmic ray energy spectrum using these values is published in [101]. The same values are used in this thesis, but a full analysis would require new simulations with updated parameters.

7.3 Data Features

Data used here are from June 2011 to May 2013 (2 years) collected by the 81-station IceTop configuration. Runs with a duration longer than 30 mins and a

stable ‘IceTopSTA5’ filter rate (good runs) are used. project. Pulses after running ‘SeededRTCleaning’ are used to reconstruct S_{125} used in the CIC method.

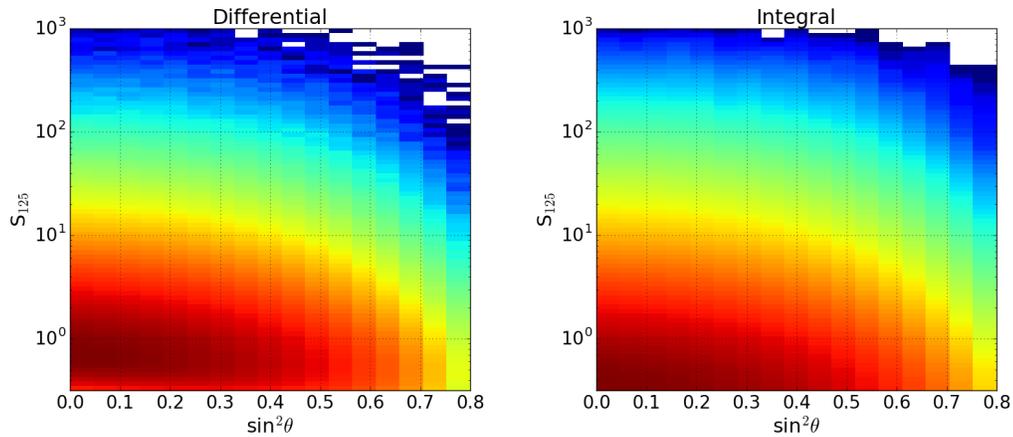


Figure 7.1: Differential (left) and Integral (right) S_{125} vs $\sin^2\theta$ spectrum. Color represents the number of events in each bin. The integral distribution at a given S_{125} has all events with S_{125} equal to or greater than that value.

- event passes IceTop_STA5_13 filter condition.
- reconstruction succeeds with ‘okay’ fit status.
- reconstructed zenith angle is less than 64° .
- shower core lies inside the IT geometry where FractionContainment is less than 0.96.
- the largest signal on a tank is greater than or equal to 6 VEM.
- $1.4 < \beta < 9.5$

Some of the features of data after all quality cuts are shown in figures 7.1, 7.2, and 7.3. Fig 7.1 shows the two-dimensional distribution of events in $\sin^2\theta$ and S_{125} bins. The left plot is the differential histogram where the number of events in each $\sin^2\theta$ and S_{125} bin is represented by the color. The right plot is the integral distribution

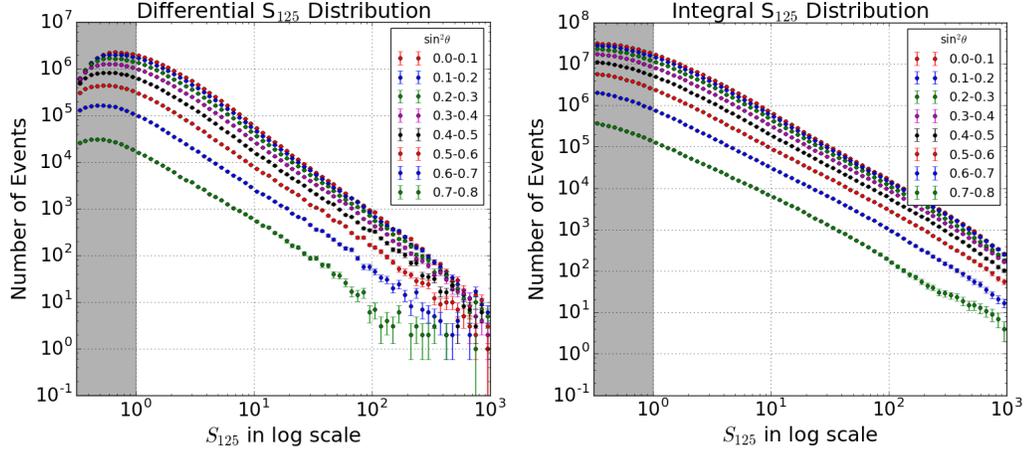


Figure 7.2: Differential (left) and integral (right) S_{125} distribution for each $\sin^2\theta$ bins. The integral distribution at a given S_{125} has all events with S_{125} equal to or greater than that value.

where the number of events in each bin is the total number of events with S_{125} equal to or greater than the S_{125} value.

The S_{125} spectrum for different $\sin^2\theta$ bins is shown in Fig 7.2. The left plot shows the differential spectrum whereas the right plot shows the integral S_{125} spectrum. The integral spectrum is the cumulative number of events with S_{125} equal to or greater than the S_{125} value. Events in the gray band are neglected because they lie significantly below the 11 VEM threshold. The reason to use 11 VEM as the threshold S_{125} is discussed in Fig 7.5.

Fig 7.3 shows the zenith distribution of events collected by IceTop in 2011 and 2012 as mentioned before. The dotted line at the center is at 27° and is approximately the mean zenith angle.

7.4 Gaisser-Hillas Type Function

The average shower size can be represented by a Gaisser-Hillas type equation as mentioned on page 239 of [129]. The shower size after convolving with the shower's

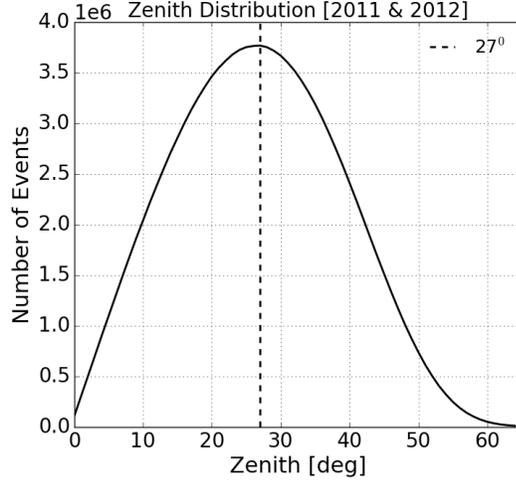


Figure 7.3: Zenith distribution of 2011 and 2012 data.

starting point distribution can be modeled by Eq 7.6.

$$S(E, X) = S_0 \frac{E}{\epsilon} \frac{X_{\max} - \lambda}{X_{\max}} e^{\left(\frac{X_{\max} - \lambda}{\lambda}\right)} \left(\frac{X}{X_{\max} - \lambda}\right)^{\frac{X_{\max}}{\lambda}} e^{\left(-\frac{X}{\lambda}\right)} \quad (7.6)$$

In this equation slant depth (X) is the variable, S_0 is the normalization factor, E is the energy of the primary, ϵ is the critical energy, X_{\max} is the slant depth where shower size is maximum, and λ is related to the interaction length of the primary particle. After taking \log_{10} of both sides of Eq. 7.6 in order to fit attenuation function, it becomes

$$\log_{10} S = p_0 + \log_{10}\left(\frac{p_1 - p_2}{p_1}\right) + \frac{p_1}{p_2} \log_{10}\left(\frac{x}{p_1 - p_2}\right) + 0.43429 \frac{p_1 - p_2 - x}{p_2} \quad (7.7)$$

where $p_0 = \log_{10}\left(\frac{S_0 E}{\epsilon}\right)$, $p_1 = X_{\max}$ and $p_2 = \lambda$. This form of the Gaisser-Hillas function is used to fit attenuation curves.

The IceTop detector is located at $\sim 692 \text{ g/cm}^2$. It will be shown later in the chapter that the shower maximum is usually above the detector. Hence, the Gaisser-Hillas function fits the tail region of the shower profile.

7.5 Attenuation Curve for Different Intensity Cuts

The S_{125} spectrum in equal $\sin^2 \theta$ bins is shown in Fig 7.2. Each S_{125} histogram must lie on top of one another if they are distributed in equal $\sin^2 \theta$ bins and if there is no attenuation. Showers coming at larger zenith angle have to pass through more matter than vertical showers. Hence, a shower with the same primary energy attenuates more at larger zenith angle resulting in a smaller signal in IceTop tanks compared to a shower with the same primary energy at smaller zenith angle. This effect can be seen in Fig 7.2. The attenuation curve is formed when these spectra are cut at constant number of events. The $\log_{10} S_{125}$ and zenith value of the points where the cut line and the spectra intersect gives the signals at various zenith bins. The top plot of Fig 7.4 shows the collection of these intersection for various cuts on the event number.

For each number of events in the integral S_{125} spectrum, the corresponding S_{125} and zenith angle are extracted. The zenith angle (θ) is converted to slant depth by $X = 692 \times \sec \theta$. The attenuation curve for 10, 35, 150, 300, 1000, 3×10^3 , 1×10^4 , 3×10^4 , 1×10^5 , and 2×10^5 number of events used for cuts are shown in the top plot of Fig 7.4. The highest curve is for 10 events and it corresponds to the highest energy. The curve gets lower in y-axis as the number of events for constant intensity cut increases. Attenuation curves show how a signal with the same primary energy detected by IceTop is attenuated due to the atmosphere. For larger slant depth the signal is attenuated more because the shower has to travel through more matter in the atmosphere.

The list of $\log_{10} S_{125}$ and its corresponding slant depths form an attenuation curve and their values for each number of events used for the constant intensity cut are shown in Appendix C Table C.1. These attenuation curves are fitted by Eq 7.7. Table C.2 shows the corresponding values of fit parameters. One of the parameters in this fit is the average shower maximum X_{\max} .

The bottom plot of Fig 7.4 shows the energy-dependent average shower maximum ($\langle X_{\max} \rangle$) from the fit using IceTop data. The energy represents the corresponding value of $\log_{10} S_{125}$ at 27° converted using Eq 7.5. The $\langle X_{\max} \rangle$ for IceTop (preliminary) are plotted on top of the left-hand side plot of Fig 8 from [130]. It

compares $\langle X_{\max} \rangle$ from different ground-based cosmic rays detectors and also shows the $\langle X_{\max} \rangle$ from simulations using three different hadronic interaction models. The $\langle X_{\max} \rangle$ follows the general trend and is close to $\langle X_{\max} \rangle$ from Tunka.

The average shower maximum increases almost linearly with energy for the same cosmic ray composition as shown in the bottom plot of Fig 7.4 for proton and iron primaries. As we have discussed before, the change in composition from lighter (proton) to heavier (iron) around the knee suggests the start of the end of galactic cosmic rays due to rigidity cutoff. Also, the change in composition from heavier to lighter around the ankle suggests the possible transition from galactic to extra-galactic cosmic rays. The $\langle X_{\max} \rangle$ as a function of energy using CIC method on IceTop data is shown in the bottom plot of Fig 7.4. Based on this result, a conclusion on the change in the cosmic ray composition around the knee and the ankle is weak due to the large error on $\langle X_{\max} \rangle$ and a further study is required.

7.6 Derivation of Cosmic Ray Flux

Values for three parameters are derived after fitting the attenuation function with Gaisser-Hillas type function. The value of these parameters depends on the number of events, i.e., each curve in the attenuation curve plot. The attenuation curve formulates how signals due to shower initiated by a primary attenuate as zenith angle increases. Using this relation, signals from any zenith angle can be transformed to the equivalent signals from 27° . The S_{125} spectrum with S_{125} calculated at 27° is shown in Fig 7.5. The energy of events must be in a region where the detector is fully efficient for the CIC method to work. Therefore events with S_{125} below 11 VEM are removed in Fig 7.5. If the CIC method is correct, all S_{125} spectra calculated at 27° should lie on top of each-other within the uncertainty, as is seen in Fig 7.5.

All $\log_{10} S_{125}$ calculated at 27° are converted to energy using Eq 7.5. An energy spectrum from the events with zenith angle less than 64° is calculated, which is then used to calculate the all-particle cosmic ray flux. A proper relation between $\log_{10} S_{125}$

to energy must be evaluated after the completion of collaboration-wide CORSIKA simulations. These simulations are not yet complete.

The left plot of Fig. 7.6 shows the $dN/d \ln E$ distribution of events from the constant intensity cut method. The right plot of Fig 7.6 shows the cosmic ray flux of IceTop 3-year energy spectrum and the result from the previous energy spectrum analysis. The main difference between these two plots is the number of events in same energy bins. The 3-year IT spectrum used three years of data whereas this analysis uses two years of data. Despite two years of data, the energy spectrum using the CIC method has more events at the higher energy bins compared to that of 3-year IT spectrum. The IceTop 3-year spectrum is calculated using events up to $\cos \theta \geq 0.8$. The data excess is due to the utilization of the CIC method which uses events up to 64° . The CIC method helps to increase statistics of events by a factor of 2 (approx.) for the energy range where the detector is fully efficient.

This chapter shows the possibility of using the CIC method for various physics goals. To complete this analysis, the analysis should be extended using simulations. The proper relation from $\log_{10} S_{125}$ to energy must be evaluated. In order to calculate the all-particle cosmic ray flux, the effective area of the detector must be calculated. Also the change in systematic uncertainty, if any, should be studied. The CIC method can depend on the properties of atmosphere like density, pressure, temperature, and humidity. As a robustness check, their effect on the results must be studied. Showers initiated by protons penetrate more than those from heavy primaries, so the angular dependence is affected by composition as well as by attenuation. The effect of composition on this method also needs to be studied.

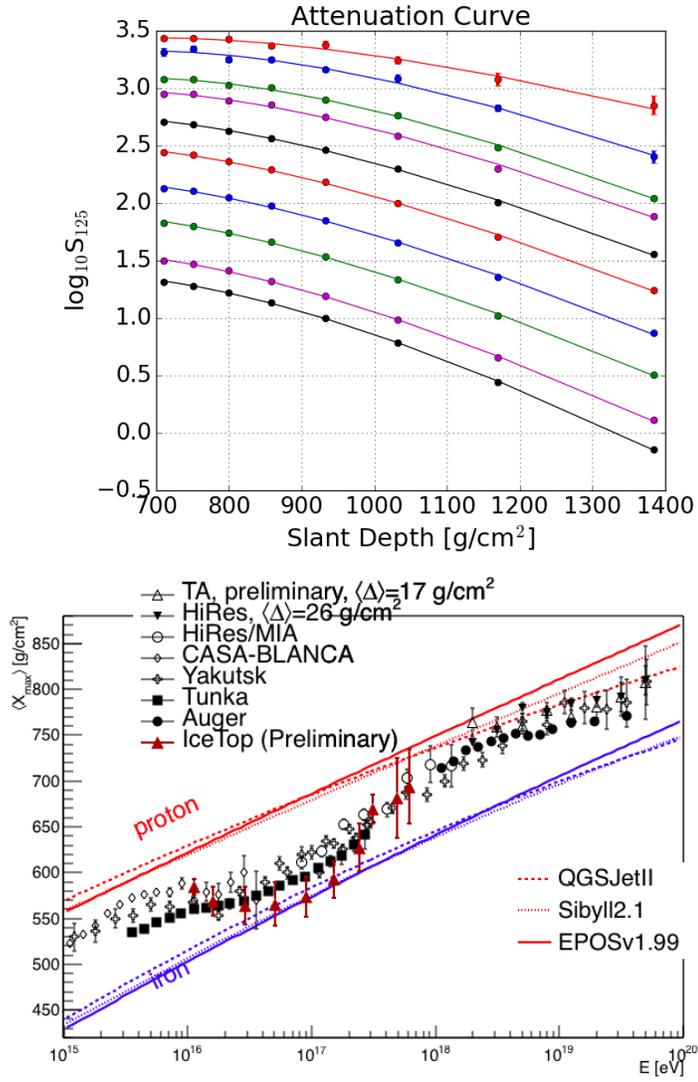


Figure 7.4: Top: Attenuation curves from integral spectra fitted with the Gaisser-Hillas function from Eq 7.7. The highest curve corresponds to the highest energy and the lowest curve corresponds to the lowest energy. Bottom: X_{\max} vs. Energy. X_{\max} is obtained after fitting attenuation curve with Eq 7.7.

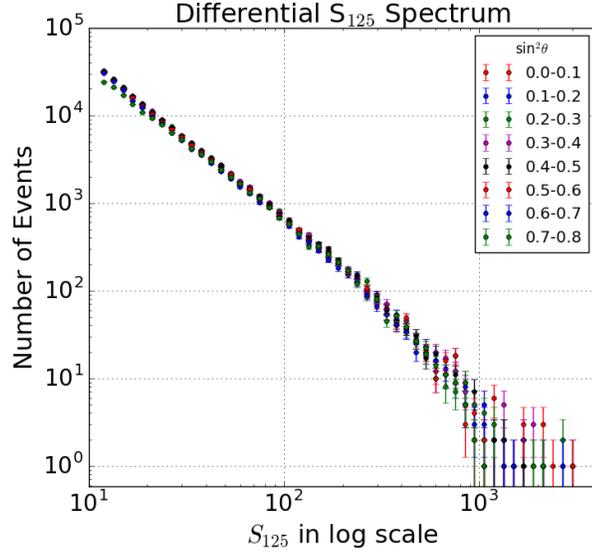


Figure 7.5: Differential spectrum of S_{125} calculated at 27° after correcting atmospheric attenuation for different $\sin^2\theta$ bins. Spectra below 11 VEM start bending away from the trend. Hence, 11 VEM is used as the threshold S_{125} value.

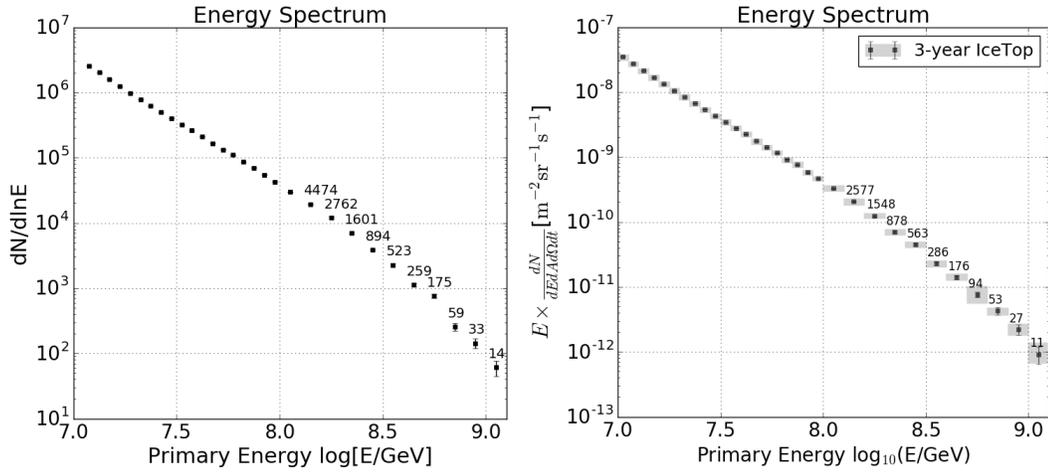


Figure 7.6: Compare the number of events in higher energy bins between energy spectrum using CIC method (left) and the 3-year IceTop energy spectrum (right). The number of events using 2 years of IceTop data and CIC method is more than the number of events using 3 years of IceTop data.

Chapter 8

RESULTS, DISCUSSION, AND FUTURE WORK

As previously mentioned, this dissertation is divided into two parts. The first part is the low energy spectrum using two stations events and the second part is the constant intensity method. This chapter describes the results from these two parts in sections 8.1 and 8.3, and discusses the results in section 8.2. Possible future work is discussed in section 8.4.

8.1 Low Energy Spectrum

The distinguishing key research of this dissertation is to lower the energy threshold of the all-particle cosmic ray energy spectrum using IceTop. Most effort went into the deployment of a new trigger and filter to select low energy events and into the development of a new reconstruction method for these hard to reconstruct events. This analysis measured the all-particle cosmic ray energy spectrum from 252 TeV to 10 PeV, lowering the energy threshold of IceTop from ~ 2 PeV to 250 TeV.

The new trigger and filter has been in operation since 20 May 2016. Data from 20 May 2016 to 28 April 2017 was used in the analysis presented here. The livetime (duration) of data used was 330.43 days. A new reconstruction method based on random forest regression was implemented. This was different from what was used for the energy spectrum in previous IceCube cosmic-ray spectrum analysis. A new set of detector simulations was generated. CORSIKA showers with Sibyll2.1 as hadronic interaction model were used. New datasets of both CORSIKA with QGSJetII-04 as hadronic interaction model and detector simulations were also generated. Several different models of the primary spectrum were used to assess the systematic uncertainty due to limited knowledge of the composition of the primary cosmic radiation.

There were three previous studies that used data from several stages of IceTop construction to measure the all-particle cosmic ray energy spectrum. The first study measured the all-particle cosmic ray spectrum from 1 PeV to 100 PeV with the partially completed 26 station IceTop configuration [98]. In that study, a maximum likelihood method was used to calculate energy proxy and Bayesian unfolding was used to calculate an unfolded energy spectrum. The second study used 40 stations of IceTop and 40 strings InIce configuration and measured composition and energy spectrum of cosmic rays with energy 1 PeV to 30 PeV [99]. The shower sampled by the surface array (mostly electromagnetic) was used in coincidence with the high-energy muon bundle from InIce. A maximum likelihood method and Neural network were used to calculate mean logarithmic mass ($\langle \ln A \rangle$) and reconstructed energy. The third study used data with 73 IceTop stations and measured the all-particle cosmic ray energy spectrum from 1.58 PeV to 1.26 EeV [100]. The latest 3-year IceTop energy spectrum analysis is an extension of IceTop-73 analysis with three years of data and 81 stations configuration [101]. The percentage of systematic uncertainties was lifted from [100] and was applied to 3-year IceTop result.

The random forest regression used in this analysis for reconstruction was a new method for energy spectrum analysis using IceTop. The containment cut in the 3-year IceTop spectrum excluded events with cores outside of the boundary created by the outermost ring of IceTop stations. In this analysis, since all shower cores lie within the IceTop geometry, a sharp geometrical containment cut was not used. The loudest station cut together with the high energy cut selected events around the infill area. This analysis used iterative Bayesian unfolding to unfold the energy spectrum which also differed from the 3-year analysis, which did not use unfolding.

Table 8.1 tabulates the result of the IceTop low energy spectrum analysis. Data from 20 May 2016 to 28 April 2017 with a livetime of 330.43 days were used. The first column is the energy bins in $\log_{10}[E/\text{GeV}]$. The second column is the number of events in reconstructed energy bins before unfolding. The total number of events in these energy bins is 7,442,086. The third column is the rate of events before

Table 8.1: Information related to all-particle cosmic ray energy spectrum using two stations events. Sibyll2.1 is the hadronic interaction model assumption. Refer to the text for detail description of each column.

$\log_{10}[E/\text{GeV}]$	N_{events}	Rate [Hz]	Unfolded Rate [Hz]	Flux [$\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}$]	Stat. Err	Sys Low	Sys High
5.4 - 5.6	3,314,095	1.1608e-1	1.2634e-1	2.1417e-5	1.7927e-8	1.5618e-6	1.3997e-6
5.6 - 5.8	2,045,186	7.1638e-2	8.2125e-2	1.0005e-5	1.2808e-8	7.7486e-7	7.3619e-7
5.8 - 6.0	1,120,728	3.9256e-2	4.6845e-2	4.9063e-6	9.3441e-9	4.2030e-7	1.9634e-7
6.0 - 6.2	527, 125	1.8463e-2	2.3261e-2	2.2929e-6	5.5339e-9	1.8326e-7	1.0734e-7
6.2 - 6.4	238,475	8.3532e-3	1.0513e-2	1.0182e-6	3.1939e-9	5.4566e-8	5.6249e-8
6.4 - 6.6	124,645	4.3660e-3	4.6427e-3	4.4639e-7	2.0603e-9	1.4168e-8	2.2057e-8
6.6 - 6.8	52,720	1.8467e-3	1.9623e-3	1.8689e-7	1.3224e-9	5.9463e-9	1.1378e-8
6.8 - 7.0	19,112	6.6945e-4	7.6403e-4	7.3287e-8	7.5675e-10	1.9554e-9	4.5247e-9

unfolding calculated by dividing the second column with livetime. The fourth column is the unfolded rate. The fifth column is the all-particle cosmic ray flux calculated using unfolded rate. The remaining columns are the statistical uncertainty, the lower systematic uncertainty, and the upper systematic uncertainty in the flux respectively. Refer to table A.3 for results assuming QGSJetII-04 as the hadronic interaction model.

8.2 Discussion

The final energy spectrum from this analysis is shown in figures 6.16 and 6.17 and is tabulated in Table 8.1. The figures show that the flux is higher than 3-year IceTop spectrum in the overlap region. Fig 8.1 shows the overlap region of the fluxes from the 3-year analysis and this analysis. These two fluxes are fitted using spline fit to calculate their percentage differences at each energy bins used in 3-year analysis. The energy spectrum from this analysis is within 8.5% of the 3-year IceTop spectrum. The total systematic uncertainty by adding individual uncertainties in quadrature for the 3-year spectrum is 9.6% at 3 PeV and 10.8% at 3 PeV [100]. Even though the flux is higher, it is within the systematic uncertainty of 3-year IceTop energy spectrum analysis. Both analyses use data collected by IceTop, so they share systematic uncertainties related to the detector. However, there are differences in this analysis, such as the treatment

of the pressure correction and the unfolding that contribute to the systematics. Other important differences are in data taking (trigger/filter) and in the use of machine learning rather than Laputop for reconstruction.

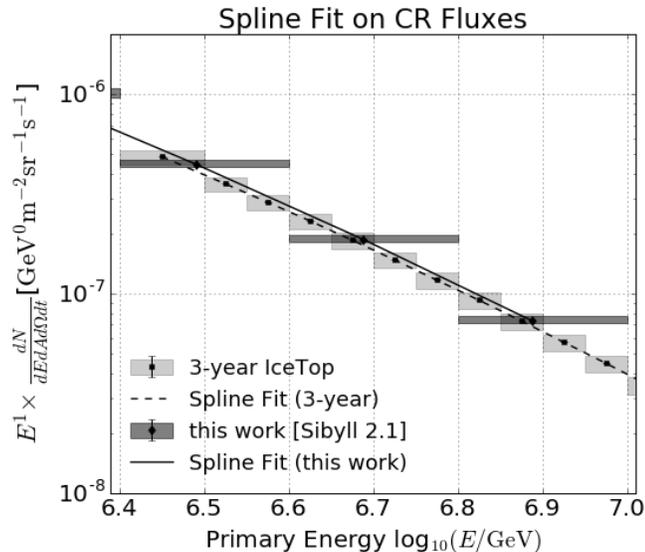


Figure 8.1: All particle cosmic ray flux showing the overlap region from 3-year IceTop analysis and this analysis. The energy spectrum from this analysis lies within the systematic uncertainty of 3-year IceTop energy spectrum.

The difference in data-taking for 3-year IceTop high energy spectrum and low energy spectrum was the use of different triggers and filters. The high energy spectrum analysis used ICE_TOP_SIMPLE_MULTIPLICITY trigger and IceTopSTA5.13 filter. The low energy spectrum analysis used ICE_TOP_VOLUME trigger and IceTop_InFill_STA2.16 filter.

The method implemented for the reconstruction of air showers for the IceTop high energy spectrum and low energy spectrum was also different. The high energy spectrum analysis uses a maximum likelihood method implemented in a software project called ‘Laputop’ to reconstruct each showers’ core position, zenith, and energy proxy (S_{125}). Simulation was then used to establish the relation from S_{125} to energy. The low energy spectrum analysis used a machine learning approach to reconstruct each showers’ core position, zenith, and energy. These differences between the two

analyses could cause a shift in energy scale and result in different cosmic ray energy spectra.

The low energy extension of the IceTop all-particle cosmic ray energy spectrum is from 250 TeV to 10 PeV. One of the distinct features of the spectrum is a bend in the knee region similar to 3-year IceTop spectrum and many other experiments. As shown in Fig 8.1, the higher energy region of IceTop low energy spectrum overlaps within a systematic uncertainty with 3-year IceTop energy spectrum. As shown in Fig 6.17, it overlaps with HAWC energy spectrum [80] within the systematic uncertainty on the lower energy region. Additionally, the energy spectrum measured in this analysis together with HAWC connects with direct measurements.

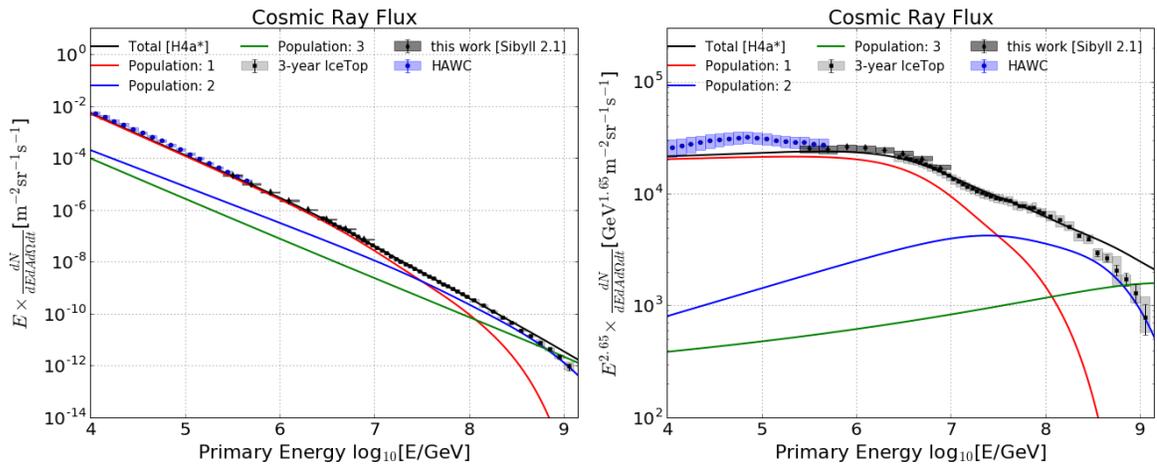


Figure 8.2: All-particle cosmic ray flux showing Peters cycle based on H4a composition model.

Fig 8.2 shows the energy spectra along with the H4a* composition model. The model and the spectrum from this analysis disagree, showing that the model does not adequately describe the data. The spectrum overlaps within systematics with the H4a model at lower energy bins but it is higher at higher energy bins. The IceTop low energy spectrum can be used as an additional information to develop a new composition model.

8.3 Constant Intensity Cut (CIC) Method

The constant intensity cut is a widely used method of calculating the cosmic ray energy spectrum. The method increases statistics of higher energy events by including air showers from higher zenith angle. In this dissertation, CIC was used to study atmospheric depth and to indirectly measure the energy-dependent average shower maximum ($\langle X_{\text{max}} \rangle$).

The signal on the ground level generated by the same particle with the same energy coming from different zenith angles is different. The shower signal attenuates more for higher zenith angle events. The attenuation of a signal detected by IceTop at different slant depths for various intensities is shown in the top plot of Fig 7.4. These attenuation curves were fitted with the Gaisser-Hillas function.

The Gaisser-Hillas function has three parameters, one of which is $\langle X_{\text{max}} \rangle$. The fit of the Gaisser-Hillas function on attenuation curves gives $\langle X_{\text{max}} \rangle$. The bottom plot of Fig 7.4 shows the $\langle X_{\text{max}} \rangle$ from the fit plotted on top of the left-hand side plot of Fig 8 in [130]. It compares $\langle X_{\text{max}} \rangle$ from different ground-based cosmic rays detectors. The figure also shows the $\langle X_{\text{max}} \rangle$ from simulations using three different hadronic interaction model. The $\langle X_{\text{max}} \rangle$ from the fit of the attenuation curve followed the general trend. It almost agrees with Tunka over all energy region.

The IceTop detector is located at an altitude of 2835 m from the sea level which corresponds to $\sim 692 \text{ g/cm}^2$. The $\langle X_{\text{max}} \rangle$ from this method ranged between 568.35 g/cm^2 and 695.73 g/cm^2 for different energies. Therefore, almost all shower maxima were located in the atmosphere before reaching the ground. In this analysis, the Gaisser-Hillas function fitted only the tail region of the longitudinal shower profile.

8.4 Future Work

The low energy spectrum analysis succeeded in lowering the energy threshold of the cosmic ray energy spectrum using IceTop to 250 TeV. To further lower the energy threshold of IceTop from 250 TeV, single station hit events with InIce information can be used. IceTop information reconstructs energy of the primary particle and

InIce information reconstructs its direction. Even for this analysis, the use of InIce information might have improved the zenith resolution from 4° .

This measured low energy spectrum is close to direct measurements where the composition of the primary is known. Therefore, a composition study can be performed in this energy region. The low energy all-particle cosmic ray spectrum could provide an extra tool for the ongoing or future study of the composition of cosmic rays. The cosmic ray energy spectrum along with its composition will help to explain the physics behind the acceleration and propagation of cosmic rays in more detail.

BIBLIOGRAPHY

- [1] T. K. Gaisser, R. Engel, and E. Resconi. *Cosmic Rays and Particle Physics*. Cambridge University Press, 2016.
- [2] G. V. Kulikov, and G. B. Khristiansen. On the Size Distribution of Extensive Atmospheric Showers. *Zhur. Eksptl'. i Teoret. Fiz.*, 35, 1958.
- [3] B. Peters. Primary cosmic radiation and extensive air showers. *Il Nuovo Cimento (1955-1965)*, 22(4):800–819, Nov 1961.
- [4] A. M. Hillas. Where do 10^{19} eV cosmic rays come from? *Nuclear Physics B - Proceedings Supplements*, 136:139 – 146, 2004.
- [5] G. Matthiae. The cosmic ray energy spectrum as measured using the Pierre Auger Observatory. *New J. Phys.*, 12:075009, 2010.
- [6] A. M. Hillas. The Origin of Ultrahigh-Energy Cosmic Rays. *Ann. Rev. Astron. Astrophys.*, 22:425–444, 1984.
- [7] K. Kotera and A. V. Olinto. The Astrophysics of Ultrahigh Energy Cosmic Rays. *Ann. Rev. Astron. Astrophys.*, 49:119–153, 2011.
- [8] V. F. Hess. Penetrating Radiation in Seven Free Ballon Flights. *Phys. Z.*, 13, 1912.
- [9] W. Kohlhörster. Messungen der durchdringenden Strahlung im Freiballon in grösseren Höhen. *Phys. Z.*, 14:1153–1156, 1913.
- [10] E. Regener and G. Pfitzer. Vertical Intensity of Cosmic Rays by Threefold Coincidences in the Stratosphere. *Nature*, 136:718–719, 1935.
- [11] A.M. Hillas. *Cosmic Rays*. Pergamon Press, 1972.
- [12] W. Baade and F. Zwicky. Remarks on Super-Novae and Cosmic Rays. *Phys. Rev.*, 46:76–77, Jul 1934.
- [13] Y. Tanaka, H. Inoue, and S. S. Holt. The X-ray astronomy satellite ASCA. *Publications of the Astronomical Society of Japan*, 46:L37–L41, Jun 1994.

- [14] K. Koyama, R. Petre, E. V. Gotthelf, U. Hwang, M. Matsuura, M. Ozaki, and S. S. Holt. Evidence for shock acceleration of high-energy electrons in the supernova remnant SN1006. *Nature*, 378:255–258, 1995.
- [15] R. Enomoto *et al.* The acceleration of cosmic-ray protons in the supernova remnant RX J1713.7-3946. *Nature*, 416:823–826, Apr 2002.
- [16] K. Murase, T. A. Thompson, B. C. Lacki, and J. F. Beacom. New Class of High-Energy Transients from Crashes of Supernova Ejecta with Massive Circumstellar Material Shells. *Phys. Rev.*, D84:043003, 2011.
- [17] B. Katz, N. Sapir, E. Waxman. X-rays, gamma-rays and neutrinos from collisionless shocks in supernova wind breakouts. *arXiv e-prints*, Jun 2011.
- [18] A. Marcowith, M. Renaud, V. Dwarkadas, and V. Tatischeff. Cosmic-ray acceleration and gamma-ray signals from radio supernovae. *Nucl. Phys. Proc. Suppl.*, 256-257:94–100, 2014.
- [19] J. A. Hinton. The Status of the H.E.S.S. project. *New Astron. Rev.*, 48:331–337, 2004.
- [20] R. Simoni, N. Maxted, M. Renaud, and J. Vink. Upper Limits on Gamma-ray Emission from Supernovae Serendipitously Observed with H.E.S.S. *Proceedings of the International Astronomical Union*, 12(S331):325–328, 2017.
- [21] J. Aleksy *et al.* (MAGIC Collaboration). Performance of the magic stereo system obtained with crab nebula data. *Astroparticle Physics*, 35(7):435 – 448, 2012.
- [22] J. Holder *et al.* The first VERITAS telescope. *Astropart. Phys.*, 25:391–401, 2006.
- [23] M. L. Ahnen *et al.* A cut-off in the TeV gamma-ray spectrum of the SNR Cassiopeia A. *Mon. Not. Roy. Astron. Soc.*, 472(3):2956–2962, 2017. [Erratum: *Mon. Not. Roy. Astron. Soc.*476,no.3,2874(2018)].
- [24] S. Kumar. A detailed study of Gamma-ray emission from Cassiopeia A using VERITAS. *PoS, ICRC2015*:760, 2016.
- [25] J. Holder. Latest Results from VERITAS: Gamma 2016. *AIP Conf. Proc.*, 1792(1):020013, 2017.
- [26] A. G. W. Cameron. Cosmic Ray Production by Vibrating Neutron Stars. *Nature*, 206:1342–1343, Jun 1965.
- [27] D. Kazanas and D. C. Ellison. Origin of ultra-high-energy gamma rays from Cygnus X-3 and related sources. *Nature*, 319:380–382, 1986.

- [28] A. Galiautdinov and D. Finkelstein. High-energy cosmic ray production by a neutron star falling into a black hole. 2017.
- [29] S. S. Kimura, K. Murase, and P. Meszaros. Super-Knee Cosmic Rays from Galactic Neutron Star Merger Remnants. *Astrophys. J.*, 866(1):51, 2018.
- [30] J. E. Gunn and J. P. Ostriker. Acceleration of High-Energy Cosmic Rays by Pulsars. *Physical Review Letters*, 22:728–731, April 1969.
- [31] T. Gold, F. G. Smith, A. W. Wolfendale, and G. D. Rochester. Pulsars and the origin of cosmic rays. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 277(1270):453–461, 1975.
- [32] W. Bednarek and M. Bartosik. Cosmic rays from galactic pulsars. *Astron. Astrophys.*, 423:405, 2004.
- [33] M. Vietri. On the acceleration of ultrahigh-energy cosmic rays in gamma-ray bursts. *Astrophys. J.*, 453:883–889, 1995.
- [34] A. Esmaili and Y. Farzan. Implications of the Pseudo-Dirac Scenario for Ultra High Energy Neutrinos from GRBs. *JCAP*, 1212:014, 2012.
- [35] E. Waxman. Cosmological gamma-ray bursts and the highest energy cosmic rays. *Phys. Rev. Lett.*, 75:386–389, 1995.
- [36] D. Biehl, D. Boncioli, A. Fedynitch, and W. Winter. Cosmic-Ray and Neutrino Emission from Gamma-Ray Bursts with a Nuclear Cascade. *Astron. Astrophys.*, 611:A101, 2018.
- [37] S. R. Kulkarni, D. A. Frail, M. H. Wieringa, R. D. Ekers, E. M. Sadler, R. M. Wark, J. L. Higdon, and E. A. Phinney. Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature*, 395:663–669, 1998.
- [38] D. Eichler, M. Livio, T. Piran, and D. N. Schramm. Nucleosynthesis, Neutrino Bursts and Gamma-Rays from Coalescing Neutron Stars. *Nature*, 340:126–128, 1989. [,682(1989)].
- [39] R. Narayan, B. Paczynski, and T. Piran. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.*, 395:L83–L86, 1992.
- [40] Y. Gao and P. M. Solomon. The Star formation rate and dense molecular gas in galaxies. *Astrophys. J.*, 606:271–290, 2004.
- [41] H. Abdalla et al. The starburst galaxy NGC 253 revisited by H.E.S.S. and Fermi-LAT. *Astron. Astrophys.*, 617:A73, 2018.

- [42] V. A. Acciari *et al.* A connection between star formation activity and cosmic rays in the starburst galaxy m82.
- [43] E. G. Berezhko. Cosmic rays from active galactic nuclei. *Astrophys. J.*, 684:L69–L71, 2008.
- [44] I. Dutan and L. I. Caramete. Ultra-High-Energy Cosmic Rays from Low-Luminosity Active Galactic Nuclei. *Astropart. Phys.*, 62:206–216, 2015.
- [45] R. J. Protheroe. High energy particles from active galactic nuclei. *Nuclear Physics B - Proceedings Supplements*, 43:229–236, 7 1995.
- [46] R. J. Protheroe. Effect of energy losses and interactions during diffusive shock acceleration: Applications to SNR, AGN and UHE cosmic rays. *Astropart. Phys.*, 21:415–431, 2004.
- [47] T. M. Venters *et al.* Energetic Particles of Cosmic Accelerators II: Active Galactic Nuclei and Gamma-ray Bursts. 2019.
- [48] M.G Aartsen *et al.* Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science*, 361(6398):147–151, 2018.
- [49] E. Massaro *et al.* The 5th edition of the roma-bzcat. a short presentation. *Astrophysics and Space Science*, 357(1):75, Apr 2015.
- [50] W. B. Atwood *et al.* (Fermi Collaboration). The Large Area Telescope on the Fermi Gamma-ray Space Telescope Mission. *The Astrophysical Journal*, 697(2):1071–1102, May 2009.
- [51] Y. T. Tanaka, S. Buson, and D. Kocevski. Fermi-LAT detection of increased gamma-ray activity of TXS 0506+056, located inside the IceCube-170922A error region, 2017.
- [52] R. Mirzoyan. First-time detection of VHE gamma rays by MAGIC from a direction consistent with the recent EHE neutrino event IceCube-170922A, 2017.
- [53] E. Orlando, G. Johannesson, I. V. Moskalenko, and T. A. Porter, and A. Strong. GALPROP cosmic-ray propagation code: recent results and updates. *Nucl. Part. Phys. Proc.*, 297-299:129–134, 2018.
- [54] V.S. Ptuskin and V. N. Zirakashvili. Limits on diffusive shock acceleration in supernova remnants in the presence of cosmic-ray streaming instability and wave dissipation. *Astron. Astrophys.*, 403:1–10, 2003.
- [55] D. Caprioli. “Espresso” Acceleration of Ultra-high-energy Cosmic Rays. *Astrophys. J.*, 811(2):L38, 2015.

- [56] E. Fermi. On the Origin of the Cosmic Radiation. *Phys. Rev.*, 75:1169–1174, Apr 1949.
- [57] R. Jansson and G. R. Farrar. The Galactic Magnetic Field. *The Astrophysical Journal Letters*, 761(1):L11, 2012.
- [58] R. Cowsik, Y. Pal, S. N. Tandon, and R. P. Verma. Steady State of Cosmic-Ray Nuclei—Their Spectral Shape and Path Length at Low Energies. *Phys. Rev.*, 158:1238–1242, Jun 1967.
- [59] R. Cowsik and L. W. Wilson. The Nested Leaky-Box Model for Galactic Cosmic Rays. *International Cosmic Ray Conference*, 2:659, Aug 1975.
- [60] V. L. Ginzburg, Ya. M. Khazan, and V. S. Ptuskin. Origin of cosmic rays: Galactic models with halo. *Astrophysics and Space Science*, 68(2):295–314, Apr 1980.
- [61] M. Aguilar *et al.* (AMS Collaboration). Precision measurement of the boron to carbon flux ratio in cosmic rays from 1.9 gv to 2.6 tv with the alpha magnetic spectrometer on the international space station. *Phys. Rev. Lett.*, 117, 2016.
- [62] T. K. Gaisser. Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio. *Astropart. Phys.*, 35:801–806, 2012.
- [63] W. Heitler. *The Quantum Theory of Radiation*. Oxford University Press, 3 edition, 1954.
- [64] J. Matthews. A Heitler model of extensive air showers. *Astropart. Phys.*, 22:387–397, 2005.
- [65] B. Rossi and K. Greisen. Cosmic-Ray Theory. *Rev. Mod. Phys.*, 13:240–309, Oct 1941.
- [66] M. Casolino *et al.* Launch of the Space experiment PAMELA. *Adv. Space Res.*, 42:455–466, 2008.
- [67] M. Aguilar *et al.* The Alpha Magnetic Spectrometer (AMS) on the International Space Station. I: Results from the test flight on the space shuttle. *Phys. Rept.*, 366:331–405, 2002. [Erratum: *Phys. Rept.* 380,97(2003)].
- [68] E.S. Seo *et al.* Cosmic-ray energetics and mass (CREAM) balloon project. *Advances in Space Research*, 33(10):1777 – 1785, 2004.
- [69] T. G. Guzik *et al.* Advanced Thin Ionization Calorimeter (ATIC) balloon experiment: instrumentation. *Proc SPIE*, pages 122–133, 10 1996.
- [70] R. Abbasi *et al.* IceTop: The surface component of IceCube. *Nucl. Instrum. Meth.*, A700:188–220, 2013.

- [71] B. Bartoli *et al.* The analog Resistive Plate Chamber detector of the ARGO-YBJ experiment. *Astropart. Phys.*, 67:47–61, 2015.
- [72] E. de la Fuente *et al.* The high altitude water čerenkov (hawc) tev gamma ray observatory. In Diego F. Torres and Olaf Reimer, editors, *Cosmic Rays in Star-Forming Environments*, pages 439–446, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [73] D. Zaborov. The HAWC observatory as a GRB detector. In *Proceedings, 4th International Fermi Symposium: Monterey, California, USA, October 28-November 2, 2012*, 2013.
- [74] T. Antoni *et al.* The cosmic-ray experiment KASCADE. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 513(3):490 – 510, 2003.
- [75] J. Huang *et al.* Measurement of high energy cosmic rays by the new Tibet hybrid experiment. *PoS, ICRC2017*:484, 2018.
- [76] N. Budnev *et al.* The Tunka detector complex: from cosmic-ray to gamma-ray astronomy. *J. Phys. Conf. Ser.*, 632(1):012034, 2015.
- [77] T. Abu-Zayyad *et al.* The surface detector array of the Telescope Array experiment. *Nucl. Instrum. Meth.*, A689:87–97, 2013.
- [78] A. Aab *et al.* The Pierre Auger Cosmic Ray Observatory. *Nucl. Instrum. Meth.*, A798:172–213, 2015.
- [79] G. Di Sciasco. Measurement of the Cosmic Ray Energy Spectrum with ARGO-YBJ. In *Proceedings, Vulcano Workshop 2014: Frontier Objects in Astrophysics and Particle Physics: Vulcano, Italy, May 18-24, 2014*, 2014.
- [80] R. Alfaro *et al.* All-particle cosmic ray energy spectrum measured by the HAWC experiment from 10 to 500 TeV. *Phys. Rev.*, D96(12):122001, 2017.
- [81] T. Antoni *et al.* KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.*, 24:1–25, 2005.
- [82] M. Amenomori *et al.* Cosmic-ray energy spectrum around the knee obtained by the Tibet experiment and future prospects. *Adv. Space Res.*, 47:629–639, 2011.
- [83] M. Amenomori *et al.* The All-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} eV to 10^{17} eV observed with the Tibet-III air-shower array. *Astrophys. J.*, 678:1165–1179, 2008.
- [84] V. Verzi, D. Ivanov, and Y. Tsunesada. Measurement of Energy Spectrum of Ultra-High Energy Cosmic Rays. *PTEP*, 2017(12):12A103, 2017.

- [85] P. Abreu *et al.* Measurement of the Cosmic Ray Energy Spectrum Using Hybrid Events of the Pierre Auger Observatory. *Eur. Phys. J. Plus*, 127:87, 2012.
- [86] B. S. Acharya *et al.* *Science with the Cherenkov Telescope Array*. WSP, 2018.
- [87] The IceCube Collaboration *et al.* Ictop: The surface component of icecube. *NIM*, 700:188–220, 2013.
- [88] The IceCube Collaboration *et al.* The icecube neutrino observatory: Instrumentation and online systems. *Journal of Instrumentation (JINST)*, 12:P03012, 2017.
- [89] C. W. Walter. The Super-Kamiokande Experiment. pages 19–43, 2008.
- [90] S. Adrian-Martinez *et al.* The prototype detection unit of the KM3NeT detector. *Eur. Phys. J.*, C76(2):54, 2016.
- [91] F. Reines and C. L. Cowan. The neutrino. *Nature*, 178:446–449, 1956.
- [92] J. Babson *et al.* Cosmic-ray muons in the deep ocean. *Phys. Rev. D*, 42:3613–3620, Dec 1990.
- [93] J. Ahrens *et al.* Results from the Antarctic Muon and Neutrino Detector Array (AMANDA). *Nucl. Phys. Proc. Suppl.*, 118:371–379, 2003. [,371(2002)].
- [94] K. Rawlins. *Measuring the Composition of Cosmic Rays with the SPASE and AMANDA Detectors*. PhD thesis, University of Wisconsin - Madison, 2001.
- [95] J.E. Dickinson *et al.* The new South Pole air shower experiment – SPASE-2. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 440(1):95–113, 2000.
- [96] M. G. Aartsen *et al.* Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector. *Science*, 342:1242856, 2013.
- [97] The IceCube Collaboration *et al.* The design and performance of icecube deepcore. *Astroparticle Physics*, 35:615–624, 2012.
- [98] The IceCube Collaboration *et al.* All-particle cosmic ray energy spectrum measured with 26 IceTop stations. *Astropart. Phys.*, 44:40–58, 2013.
- [99] The IceCube Collaboration *et al.* Cosmic ray composition and energy spectrum from 1-30 PeV using the 40-string configuration of IceTop and IceCube. *Astroparticle Physics*, 42:15–32, Feb 2013.
- [100] M. G. Aartsen *et al.* Measurement of the cosmic ray energy spectrum with IceTop-73. *Phys. Rev.*, D88(4):042004, 2013.

- [101] K. Rawlins. Cosmic ray spectrum and composition from three years of IceTop and IceCube. *J. Phys. Conf. Ser.*, 718(5):052033, 2016.
- [102] IceCube Collaboration, in preparation. Cosmic Ray Spectrum and Mass Composition from IceTop and IceCube. *TeVPA2017*, 2018.
- [103] T. K. Gaisser, T. Stanev, T. Waldenmaier, and X. Bai. IceTop/IceCube coincidences. In *Proceedings, 30th International Cosmic Ray Conference (ICRC 2007): Merida, Yucatan, Mexico, July 3-11, 2007*, volume 5, pages 1209–1212, 2007.
- [104] IceCube Collaboration, in preparation. Search for PeV Gamma-Ray Point Sources with IceCube. *ICRC2017*, 301, 2018.
- [105] R. Abbasi *et al.* Solar Energetic Particle Spectrum on 13 December 2006 Determined by IceTop. *Astrophys. J.*, 689:L65–L68, 2008.
- [106] Hamamatsu Photonics K.K. Photomultiplier Tubes; Basics and Applications, 2007.
- [107] D. E. Groom, and N. V. Mokhov, and S. I. Striganov. Muon Stopping Power and Range Tables 10 MeV-100 TeV. *Atomic Data and Nuclear Data Tables*, 78:183–356, Jul 2001.
- [108] A. Ismail. Measurement of the cosmic ray energy spectrum between 500 tev and 100 pev with icetop, 2015.
- [109] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. 1998.
- [110] S. Agostinelli. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [111] E. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev. Cosmic ray interaction event generator SIBYLL 2.1. *Phys. Rev. D*, 80:094003, Nov 2009.
- [112] S. Ostapchenko. Air shower development: impact of the LHC data. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011): Beijing, China, August 11-18, 2011*, volume 2, page 71, 2011.
- [113] T. K. Gaisser, T. Stanev, S. Tilav. Cosmic Ray Energy Spectrum from Measurements of Air Showers. *Front. Phys.(Beijing)*, 8:748–758, 2013.
- [114] H. Dembinski *et al.* Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV. *PoS, ICRC2017*:533, 2018.
- [115] J. R. Hoerandel. On the knee in the energy spectrum of cosmic rays. *Astropart. Phys.*, 19:193–220, 2003.

- [116] S. Ostapchenko. High Energy Cosmic Ray Interactions: An Overview. *J. Phys. Conf. Ser.*, 60:167–170, 2007.
- [117] A. Fedynitch, J. Becker Tjus, and P. Desiati. Influence of hadronic interaction models and the cosmic ray spectrum on the high energy atmospheric muon and neutrino flux. *Phys. Rev.*, D86:114024, 2012.
- [118] G. James, D. Witten, T. Hastie, Trevor, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.
- [119] L . Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [120] X. Meng *et al.* MLlib: Machine Learning in Apache Spark. *J. Mach. Learn. Res.*, 17(1):1235–1241, January 2016.
- [121] M. Zaharia *et al.* Resilient Distributed Datasets: A Fault-tolerant Abstraction for In-memory Cluster Computing. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, NSDI’12, pages 2–2. USENIX Association, 2012.
- [122] F. Pedregosa *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, Nov 2011.
- [123] G. D’Agostini. Improved iterative bayesian unfolding. *physics.data-an*, 2010.
- [124] J. Bourbeau and Z. Hampel-Arias. Pyunfold: A python package for iterative unfolding. *physics.data-an*, 2018.
- [125] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*, 1945.
- [126] G. Cowan. *Statistical Data Analysis*. Oxford University Press, 1998.
- [127] W.D. Apel *et al.* The KASCADE-Grande energy spectrum of cosmic rays and the role of hadronic interaction models.
- [128] J. Alvarez-Muniz, R. Engel, T. K. Gaisser, J. A. Ortiz, and T. Stanev. Atmospheric shower fluctuations and the constant intensity cut method. *Phys. Rev.*, D66:123004, 2002.
- [129] T. K. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University Press, 1 edition, 1990.
- [130] K. Kampert and M. Unger. Measurements of the Cosmic Ray Composition with Air Shower Experiments. *Astropart. Phys.*, 35:660–678, 2012.

Acronyms

A_{eff} Effective area

AGN Active Galactic Nuclei

AIRS Atmospheric Infrared Sounder

AMRC Antarctic Meteorological Research Center

ATWD Analog Transient Waveform Digitizer

DAQ data acquisition

DOM Digital Optical Module

DOR DOM readout card

eV electron-volt

FPGA Field Programmable Gate Array

GeV giga-electronvolt

GRB Gamma Ray Burst

HAWC High Altitude Water Cherenkov

ICL IceCube Lab

pc parsec

pe photo-electron

PMT photo-multiplier tube

pulsar Pulsating Radio Star

RAPCal Reciprocal Active Pulsing Calibration

RSS Residual Sum of Squares

VEM Vertical Equivalent Muon

Appendix A
SPECTRUM RELATED

Layer i	Altitude h (km)	a_i (g/cm ²)	b_i (g/cm ²)	c_i (cm)
1	0 ... 4	-163.331	1183.70	875221.
2	4 ... 10	-65.3713	1108.06	753213.
3	10 ... 40	0.402903	1424.02	545846.
4	40 ... 100	-0.000479198	207.595	793043.
5	> 100	0.00188667	1	$5.9787908 \cdot 10^9$

Figure A.1: Parameters of South pole atmosphere for Jul. 01, 1997 (MSIS-90-E) after fitting with Eq. 4.1. This table is copied from CORSIKA user-guide.

Table A.1: Parameter after fitting effective area with sigmoid type function in Eq 6.6. The first row is Aeff parameters for events with $\cos\theta \geq 0.9$ assuming H4a composition model. The second row is Aeff parameters assuming all Proton composition. The third row is Aeff parameters assuming all Iron composition.

	p_0	p_1	p_2
H4a [$\cos\theta \geq 0.9$]	3.58850483e+04	5.22717015	5.4472083
Proton [$\cos\theta \geq 0.9$]	3.59613147e+04	4.58304647	5.34671533
Iron [$\cos\theta \geq 0.9$]	3.61815496e+04	5.91349511	5.58050636
H4a [$0.8 \leq \cos\theta \leq 0.9$]	3.62042019e+04	5.01271692	5.57632802

Table A.2: Information related to all-particle cosmic ray energy spectrum using two-station events and assuming Sibyll 2.1 as the hadronic interaction model.

log[E/GeV]	N_{events}	Rate [Hz]	Unfolded Rate [Hz]	Flux [m ⁻² s ⁻¹ sr ⁻¹]	Stat. Err	Sys Low	Sys High
5.4 - 5.6	3,314,095	1.1608e-1	1.2634e-1	2.1417e-5	1.7927e-8	1.5618e-6	1.3997e-6
5.6 - 5.8	2,045,186	7.1638e-2	8.2125e-2	1.0005e-5	1.2808e-8	7.7486e-7	7.3619e-7
5.8 - 6.0	1,120,728	3.9256e-2	4.6845e-2	4.9063e-6	9.3441e-9	4.2030e-7	1.9634e-7
6.0 - 6.2	527, 125	1.8463e-2	2.3261e-2	2.2929e-6	5.5339e-9	1.8326e-7	1.0734e-7
6.2 - 6.4	238,475	8.3532e-3	1.0513e-2	1.0182e-6	3.1939e-9	5.4566e-8	5.6249e-8
6.4 - 6.6	124,645	4.3660e-3	4.6427e-3	4.4639e-7	2.0603e-9	1.4168e-8	2.2057e-8
6.6 - 6.8	52,720	1.8467e-3	1.9623e-3	1.8689e-7	1.3224e-9	5.9463e-9	1.1378e-8
6.8 - 7.0	19,112	6.6945e-4	7.6403e-4	7.3287e-8	7.5675e-10	1.9554e-9	4.5247e-9

Table A.3: Information related to all-particle cosmic ray energy spectrum using two-station events and assuming QGSJetII-04 as the hadronic interaction model.

log[E/GeV]	N_{events}	Rate [Hz]	Unfolded Rate [Hz]	Flux [m ⁻² s ⁻¹ sr ⁻¹]	Stat. Err	Sys Low	Sys High
5.4 - 5.6	3,314,095	1.1608e-1	1.1894e-1	2.2632e-5	1.4461e-8	3.8603e-6	2.2162e-6
5.6 - 5.8	2,045,186	7.1638e-2	8.0321e-2	9.6311e-6	7.5422e-9	1.2275e-6	9.1223e-7
5.8 - 6.0	1,120,728	3.9256e-2	4.8040e-2	4.7506e-6	5.7301e-9	6.8984e-7	4.2067e-7
6.0 - 6.2	527, 125	1.8463e-2	2.4486e-2	2.2660e-6	4.3013e-9	3.6153e-7	2.0890e-7
6.2 - 6.4	238,475	8.3532e-3	1.0851e-2	9.8810e-7	2.3529e-9	1.2659e-7	7.7093e-8
6.4 - 6.6	124,645	4.3660e-3	4.2625e-3	3.8602e-7	1.2589e-9	3.5333e-8	2.9719e-8
6.6 - 6.8	52,720	1.8467e-3	1.7742e-3	1.5931e-7	8.5679e-10	1.2446e-9	1.4503e-8
6.8 - 7.0	19,112	6.6945e-4	7.2121e-4	6.5249e-8	5.8469e-10	4.5713e-9	5.7654e-9

Table A.4: Quality of reconstruction. The first row shows the core resolution in meter. The second row shows the zenith resolution in degree. The third row shows the energy resolution. All these resolutions are one sigma of true minus reconstructed value.

log[E/GeV]	5.4-5.6	5.6-5.8	5.8-6.0	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
Core [m]	15.62	13.85	12.03	9.76	8.45	7.76	6.95	6.22
Zenith [deg]	3.95	3.47	2.87	2.51	1.94	1.95	1.62	1.46
log[E _{true} /E _{reco}]	0.26	0.24	0.20	0.16	0.14	0.12	0.10	0.09

Appendix B
COMPOSITION RELATED

Table B.1: Parameters for modified Polygonato composition model. An exponential cutoff in rigidity of population 1 is performed here compared to the original Polygonato composition model.

Rc	gamma	p	He	CNO	Mg-Si	Fe
γ for Pop.1	-	1.71	1.64	1.67	1.70	1.55
Pop.1: 4.49 PV	see line 1	11800	4750	3860	3120	1080
Pop.2: 30 PV	1.4	11.8	11.8	7.88	7.88	7.88
Pop.3: 2 EV	1.4	1.7	1.7	1.14	1.14	1.14

Appendix C
CIC RELATED

Table C.1: $\log_{10} S_{125}$ values used in attenuation curve for each curve generated for some constant number of events cut. $\log_{10} S_{125}$ values for the corresponding slant depth (g/cm²) (X) is tabulated below.

N_{event}/X	709.98	750.58	799.058	858.328	933.098	1031.578	1169.69	1384.
10	3.4287	3.4354	3.4250	3.3670	3.3750	3.2419	3.0750	2.8487
35	3.3097	3.3420	3.2496	3.2456	3.1605	3.0806	2.8250	2.4012
150	3.0761	3.0750	3.0250	3.0014	2.8947	2.7619	2.4808	2.0419
300	2.9499	2.9443	2.8892	2.8550	2.7438	2.5846	2.2996	1.8824
1e3	2.7009	2.6828	2.6266	2.5634	2.4620	2.2935	2.0037	1.5541
3e3	2.4359	2.4161	2.3578	2.2867	2.1794	1.9949	1.7035	1.2424
1e4	2.1240	2.1037	2.0442	1.9716	1.8471	1.6538	1.3501	0.8655
3e4	1.8260	1.7980	1.7424	1.6611	1.5322	1.3337	1.0152	0.5068
1e5	1.4969	1.4645	1.4081	1.3202	1.1870	0.9799	0.6516	0.1096
2e5	1.3136	1.2786	1.2214	1.1297	0.9934	0.7804	0.4429	-0.1442

Table C.2: Parameter after fitting attenuation curve with Gaisser-Hillas function. p_1 represents the mean shower maximum X_{max} .

N_{event}	10	35	150	300	1000	3e3	1e4	3e4	1e5	2e5
p_0	3.48	3.36	3.12	3.02	2.79	2.55	2.25	1.95	1.62	1.41
p_1	693.95	681.95	669.36	627.42	592.84	573.49	565.56	563.94	568.44	584.77
Error (p_1)	40.18	43.59	15.95	25.96	19.94	21.85	23.54	20.71	15.88	8.76
p_2	148.02	104.92	94.42	101.93	103.66	104.24	100.37	96.37	90.96	83.74