ENERGY ANISOTROPIES OF PROTON-LIKE ULTRA-HIGH ENERGY COSMIC RAYS

Jon Paul Lundquist

Thesis Defense
September 22, 2017
INTRODUCTION

PART ONE

- Overview of Ultra-high Energy Cosmic Rays (UHECR)
  - Sources and propagation
  - Previous Results
    - Anisotropy
    - Energy Spectrum
    - Composition
  - Extensive Air Showers
  - Telescope Array Project and UHECR Detection

PART TWO

- Anisotropy Studies
  - Kernel Density Estimation Hotspot Analysis (not enough time)
  - Energy Spectrum Anisotropy
  - Energy-Distance Correlation
  - Hot/Coldspot Summary Analysis

PART THREE

- Composition Study
  - Pattern Recognition Analysis
  - Composition
    - L-test and the Shift plot
GOALS

- Remove model dependencies – assumptions and parameters – whenever possible.
  - Requires the development of new statistical methods.

- Anisotropy – magnetic deflection as a signature of a source instead of a confounding variable

- Combine UHECR energy, anisotropy, and composition into one picture.
PART ONE
OVERVIEW OF ULTRA-HIGH ENERGY COSMIC RAYS
SOURCES AND PROPAGATION

**SOURCES**

- **Intergalactic EM Field (1-10 nG)?**

- **Greisen–Zatsepin–Kuzmin limit**
  
  \[ E < \frac{E_{\text{max}}}{10^{18} \text{eV}} \approx \frac{1}{2} \beta \cdot Z \cdot \frac{B}{\mu G} \cdot \frac{L}{\text{kpc}} \]

- **Cosmic Microwave Background**

**RELATES:** energy, sources, composition

- **Larmor radius energy limitation**

\[ \delta = \frac{S}{R_{\text{Larmor}}} = 0.5^\circ Z \frac{L \cdot B}{\text{kpc} \cdot \mu G} \cdot 10^{20} \frac{E}{\text{eV}} \]

**Intragalactic EM Field**

**Total Deflection**

\[ <\delta> \approx \sim 10^\circ \quad \Rightarrow \delta \lesssim 50^\circ \text{ for } E = 10^{20} \]
PREVIOUS RESULTS - ANISOTROPY

Hotspot near the supergalactic plane:

- Ursa Major cluster (20 Mpc from Earth)
- Coma cluster (90 Mpc)
- Virgo cluster (20 Mpc)

The angular distance to the supergalactic plane is $\sim 17^\circ$.

3.4$\sigma$ significance of the anisotropy observation

"A Monte Carlo Bayesian Search for the Plausible Source of the Telescope Array Hotspot" He, H.N. et al.

FIG. 2: The 19 events at the hotspot in the equatorial coordinates are denoted by filled circles (red: $E < 75$EeV; blue: $E > 75$EeV). Reconstructed positions of shifted sources for two groups of the hotspot events are denoted by the open squares; the errors are shown by ellipses of the corresponding color.
PREVIOUS RESULTS – ENERGY SPECTRUM

Good Agreement Between TA Detectors

Good Agreement Between Experiments

“Ankle” is approximately the end of galactic sources
Published 5 year data result from this thesis work
- proton-dominant *depending on model*

- PAO reports heavier at higher energies
  - Data of TA and PAO agree
  - Result of different simulations or North/South anisotropy?
EXTENSIVE AIR SHOWERS (EAS)

- Muons created early in shower -- charged Pion decays.
- Muons and neutrinos are “missing energy”

EM Cascade is the largest contribution.

- Collisions result in Pions, Kaons (~8%), and Nucleons (~4-5%)
- 1/3 of the Pions ($\pi^0$) decay into two photons and contribute to EM cascade in each generation.
- Kaons contribute ~8% to EM cascade in later generations.

UHECR studied indirectly using extensive air shower
Radiation length, $\lambda$ (36.5 g/cm$^2$ in air), is about the same for pair production and Bremsstrahlung radiation.

$X_{\text{max}}$: shower depth with maximum particles. The shower then decreases in size due to ionization losses.
INDIRECT DETECTION

• Fluorescence from charged particle excitation of nitrogen.

• Charged particles activate scintillator plastic. Picked up by photomultiplier tubes.
CORSIKA Simulated Air Shower $10^{15}$ eV $45^\circ$ inclination
Red – $e^\pm$, $\gamma$
Green – $\mu^\pm$
Blue – Hadrons ($\pi^{0/+/-}$, $K^{0/+/-}$, p, n)

Reconstructions used from previous works

Fluorescence Detector
Surface Detector
**RECONSTRUCTION**

Viewing angle converted to slant depth using atmospheric profile

Fit to GH profile and search for the reconstructed Monte Carlo event that matches the data

\[
N(X) = N_{max} \frac{(X - X_0)^{X_{max} - X_0}}{\lambda} \exp \left( \frac{X_{max} - X}{\lambda} \right)
\]

Take energy from MC event which minimizes

\[
\chi^2_{\text{Profile}} = \sum_i \frac{1}{\sigma_i^2} \left( S_i^{(m)} - S_i^{(p)} \right)^2
\]

\( S_i \) are measured and predicted tube signals
PART TWO
ANISOTROPY STUDIES
DATA SUMMARY

Data:
- 7 years surface detector data \textit{(from ICRC hotspot)}.
  \# Detector. \(\geq 4\), Zenith angle \(< 55^\circ\), Pointing Error \(< 10^\circ\)

  Additional cuts \textit{(due to lower energy)}:
  - Pointing direction error \(< 5^\circ\), boundary \(> 1.2 \text{ km}\), Lateral fit \(\chi^2 < 10\)
  - \(E \geq 10^{19.0} \text{ eV} - 3027 \text{ events}\)
ISOTROPIC MONTE-CARLO COMPARISON

- $\sin(\theta) \times \cos(\theta)$ – Zenith distribution from detector geometry
- Flat Azimuth angle distribution.
- On-time simulated – sampling 250,000 event times ($E > 17.7$ EeV).
- Energy sampled from reconstructed HiRes spectrum.

$(E > 20$ EeV, $p = 0.48)$
ENERGY SPECTRUM ANISOTROPY

Is there a location on the sky which has a significantly different overall spectrum? Signature of sources, magnetic deflection or both.
Period: 2008 May – 2015 May

Cuts:
- # of used detectors ≥ 4
- Zenith angle < 55°
- Pointing Error < 10°
- Energy Threshold ≥ 57EeV

Resulting Data: 109 events

20° binning

3.4σ post-trial significance

Max. significance 5.1σ
148.5° R.A, 44.5° Dec.
(17° from SGP)

Energy distribution shows an overall deficit of events

Tighter Cuts, 20° bin

\( \chi^2 / \text{dof} = 56.9 / 19 \)

Inside (On): Mean 19.2, RMS 0.26

Outside (Off): Mean 19.2, RMS 0.2
METHOD
ESTIMATED BACKGROUND – EQUAL EXPOSURE

- Likelihood and $\chi^2$ tests are sample size biased
  - Need to control statistics

- Equal exposure binning samples the sky equally.
  - “On” exposure such that bin size average = 15°, 20°, 25°, 30°

- $30^\circ < \text{bin}, \ E \geq 10^{19.2} \text{ eV}$

- Maximum pre-trial significance for mean bin size of 30°
- Energy threshold scanned - $10^{19.0}, 10^{19.1}, 10^{19.2}, 10^{19.3} \text{ eV}$.

\[ \alpha = \frac{N_{on}^{MC}}{N_{off}^{MC}} = \text{constant} \]
OVERSAMPLING GRID

0.5°x0.5° in RA and Decl.

Changing sampling -- declination bias

0.3° +/- 0.2° Opening Angle
Median 0.32°

Histograms of closest grid distance


Used here
0.5°x0.5° in Opening Angle*

Sky is sampled equally

0.52° +/- 0.03°
Median 0.50°
POISSON LIKELIHOOD GOODNESS-OF-FIT

• Compare energy distribution “On” (inside) to “Off” (outside)
  • “Off” Normalized to $N_{bg}$ (expectation)
  
  • Energy bins of 0.05 log$_{10}$($E$/eV)
    • Less than mean energy resolution

\[ \sum n_k^{bg} = N_{bg} = \alpha N_{off}^{data} \]

• $n_{on}$ # data in bin
• $n_{bg}$ expectation
• Degrees of freedom:
  • # bins
  • +1 for fluctuating background
  • +1 for variable number of bins

Test Used Previously by T.A. In:

• Good reference http://www.fysik.su.se/~conrad/James/james.5.gof.pdf or Particle Data Group book
RESULT
ENERGY SPECTRUM ANISOTROPY – 30° <BIN>

- σ deviation — “On” data compared to “Off” data

- Maximum: 6.17σ
- 138.8° R.A., 44.8° Decl.
- Bin size: 28.43°
- # Events: 147
- 6.8° from “hotspot”
ENGLISH COMPARISON – MAX. LOCAL SIGMA

- Max. local $\sigma$ (6.17) location — 138.8° R.A., 44.8° Decl.
- 28.43° radius cap bin
- $E \geq 10^{19.2}$ eV
- Expected Background: $N_{bg} = 166.2$

Bin Chi Squares

$$\chi_k^2 \approx 2n_{on} \log \frac{n_{on}}{n_{bg}} + n_{bg} - n_{on}$$
GLOBAL SIGNIFICANCE

• Count simulations with $\sigma \geq 6.17$

• MC TEST Penalties
  • Bin scan - 15°, 20°, 25°, 30° average bin sizes
    • Not enough events inside bins less than 15°
    • Not enough events outside bins greater than 30°

  • Energy threshold scan - $10^{19.0}, 10^{19.1}, 10^{19.2}, 10^{19.3}$ eV.
    • Not enough events for cuts > $10^{19.3}$ eV

• Max. $\sigma$ of 4*4 = 16 is counted as 1 MC.

**Result:** from 2,500,000 sets of 16 maps **232 passed for 3.74$\sigma_{global}$**

*One sided probability with 16 times scan penalty*
SPECTRUM ANISOTROPY – GLOBAL SIGNIFICANCE

MC trials maximum distribution

Local sigma to Global post-trial sigma
SPECTRUM ANISOTROPY – GLOBAL SIGNIFICANCE

- 138.8° R.A., 44.8° Decl.
- Local sigma: 6.17σ
- Global sigma: 3.74σ

Rough estimate of radius: 1659 grid points σ>0.7. sqrt((1659*0.5)/pi) ≈15°
**INTEGRAL DAY SIGNIFICANCE**

- $\sigma_{local}$ at 7 year max location $\sim +1$ $\sigma$/year
  - Linear correlation 0.989

- Maximum $\sigma_{local}$ on map
  - Linear correlation 0.976

- Blue line is linear fit
POSSIBLE CAUSES

• Possible source:
  • M82 starburst galaxy most likely source
    • “A Monte Carlo Bayesian Search for the Plausible Source of the Telescope Array Hotspot”
      https://arxiv.org/abs/1411.5273
    • “Ultra-high-energy-cosmic-ray hotspots from tidal disruption events”
      https://arxiv.org/abs/1512.04959

• Possible magnetic field:
  • Supergalactic magnetic sheet increases post-GZK flux (E > 50 EeV) and deflects (E < 50 EeV)
    • “The supergalactic structure and the origin of the highest energy cosmic rays”
    • “Cosmic Magnetic Fields in Large Scale Filaments and Sheets”
ENERGY SPECTRUM ANISOTROPY
CONCLUSION

• There is a $3.74\sigma$ Energy Spectrum Anisotropy ($E \geq 10^{19.2} \text{ eV}$) at $138.8^\circ$ R.A., $44.8^\circ$ Decl.
  • Deficit at low energies and excess at high energies
  • It has been increasing in significance every year.
• Evidence of magnetic deflection of UHECR
ENERGY-DISTANCE CORRELATION

Is there a direct signature of magnetic deflection?
**GOAL**

- Anisotropy search with fewest assumptions
  - Magnetic fields deflect low energy more than high energy.
  - Single dominant source

- **No assumptions for:**
  - source distribution
  - event composition
  - magnetic field configurations.

\[ E \geq 20 \text{ EeV} \]
SOME PREVIOUS STUDIES

Most similar to this analysis


• Parameters:
  • 20 EeV threshold – **USED IN THIS ANALYSIS**
  • Lots of other parameters:
    • Linear correlations with inverse energy
    • directional with limit on transverse spread
    • 20 deg. distance limit
    • one event E>45 EeV required
    • limit on minimum correlation
    • There are a number of other hidden parameters as well…

“…there is no significant evidence for the existence of correlated multiplets in the present data set.”

SOME PREVIOUS STUDIES

• Search for patterns by combining cosmic-ray energy and arrival directions at the Pierre Auger Observatory.

  Parameters:
  • 5 EeV cut
  • Lots of other parameters:
    • Number of iterations
    • Cone size
    • 60 EeV events as center of cones.

  “...using observables sensitive to patterns characteristic for deflections in cosmic magnetic fields. No such patterns have been found within this analysis.”

METHOD
ENERGY-DISTANCE RANKED CORRELATION

For each event \((i)\) Kendall’s \(\tau\) correlation
\[
F_i \left[ E_j (E_j > E_i), \theta_{ij} (E_j > E_i) \right]
\]

\[
\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2} n (n - 1)}
\]

- **No binning**
- **Each event becomes a test point.** Test parameters (location and energy cut) decided by the data. Removes free parameters.

B is a test point
C is an event with Energy >= B
a is the Opening Angle

Robust against outliers

Linear Corr: 0.903
Outlier decreases corr: ~0.02

Rank Corr: 0.9994
Outlier decreases corr: 0.0006

\[ \Delta \sigma = \arccos(n_1 \cdot n_2) \]
RESULT
Each test point \( (i) \) calculate Kendall’s \( \tau \) correlation \( F_i[E_j(E_j > E_i), \theta_{ij}(E_j > E_i)] \)

- **Negative Correlation:** Energy Increases (decreases) \( \rightarrow \) Angle decreases (increases)
- **Positive Correlation:** Energy Increases (decreases) \( \rightarrow \) Angle Increases (decreases)

- Size proportional to \( 1/p \)-value
- Color is Distance/Energy correlation

*Each Test Point sample size is different*

Events with \( E \geq 20 \text{ EeV} \)

2 MOST SIGNIFICANT POINTS

R.A. 119.6, Dec. 59.2
E \geq 75.0 \text{ EeV}

<table>
<thead>
<tr>
<th>tau</th>
<th>p-val</th>
<th>$\sigma$-local</th>
<th>$\sigma$-global</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.452</td>
<td>0.000927</td>
<td>2.9</td>
<td>2.46</td>
</tr>
</tbody>
</table>

R.A. 154.6, Dec. 54.6
E \geq 41.2 \text{ EeV}

<table>
<thead>
<tr>
<th>tau</th>
<th>p-val</th>
<th>$\sigma$-local</th>
<th>$\sigma$-global</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.188</td>
<td>0.000167</td>
<td>3.4</td>
<td>2.33</td>
</tr>
</tbody>
</table>
Localization of Effect with Unbinned Data

- For each test point \( i \) PARTIAL LINEAR correlation of all test points \( |\tau|'s - F_i [|\tau_j|, \theta_{ij}] \)
  - Account for differing sample size by controlling for p-value

- **Negative Correlation:** Correlations decrease from that point. A “source.”
- **Positive Correlation:** Correlations increase from that point.

- Size proportional to \( 1/p \), Color is \( \rho |\tau|, p \)

Correlation of correlations assumes single source

- Circular Mean Weighted by \( \tau \)'s \( 1/p \)
  - 126.2° R.A., 48.4° Dec.

- Maximum \( |\rho |\tau|, p| \):
  - 125.9° R.A., 49.7° Dec.

*Each test point sample size is the same*
NOT A MEASURE OF DENSITY

- Data declination is subtracted and folded up toward top of FOV to see how location is tracked
- Steps of -5 deg. Left to right, top to bottom.
CORRELATION ANISOTROPY SIGNIFICANCE TEST

Single parameter search in MC – use $\sigma_{\rho|\tau|,p}$

- $\sigma_{\rho|\tau|,p} = 6.47\sigma$, $\rho_{|\tau|,p} = -0.22$
- 125.9° R.A., 49.7° Dec. – 9.4° from Energy Spectrum Anisotropy maximum

Count MC (or energy scrambled data) with $\sigma \geq \sigma_{data}$ for $\rho_{|\tau|,p} < 0$

Result: 556 (or 521) counts out of 1,500,000 - that’s $3.37\sigma$
Count MC (or energy scrambled data) with $\sigma \geq \sigma_{\text{data}}$ for $\rho_{|\tau|,p} < 0$

Result: 556 (or 521) counts out of 1,500,000 - that’s 3.37\sigma
INTEGRAL DAY SIGNIFICANCE

8 year estimate $\sim 4\sigma_{\text{global}}$

- $\sigma_{\text{local}}$ at 7 year max location — +1 $\sigma$/year
  - Linear correlation 0.910 (0.944 after 5th year)

- Maximum $\sigma_{\text{local}}$ on map
  - Linear correlation 0.905 (0.935 after 5th year)
ENERGY-DISTANCE CORRELATION

CONCLUSION

- There is a 3.37σ Energy/Distance Correlation Anisotropy (E≥10^{19.3} eV) located at 125.9° R.A., 49.7° Dec.
  - It has increased in significance 6 out of 7 years.

- Direct evidence of magnetic deflection of UHECR
COMBINED MEASURE OF ENERGY SPECTRUM ANISOTROPY AND ENERGY/POSITION CORRELATION

- Energy Spectrum Anisotropy significance: $3.74\sigma$
- Energy-Distance Correlation Anisotropy significance: $3.37\sigma$
- Stouffer’s Method combined significance: $5.03\sigma$

NEXT UP:
- Hot/Coldspot Anisotropy (a result of Spectrum AND Position Correlation): $5.4\sigma$
HOT/COLDSPOT SUMMARY ANALYSIS

Is there a direct signature of magnetic deflection?
METHOD
LI-MA SIGNIFICANCE

- Compare N events “On” (inside) to expectation – How significant is the excess or deficit?
- Derived by Poisson likelihood ratio and approximation to $\chi^2$ (like the Poisson Likelihood GOF)

$$S = \text{sign}(N_{on} - N_{bg}) \sqrt{2} \left\{ N_{on} \ln \left[ \frac{1 + \alpha}{\alpha} \left( \frac{N_{on}}{N_{on} + N_{off}} \right) \right] + N_{off} \ln \left[ (1 + \alpha) \left( \frac{N_{off}}{N_{on} + N_{off}} \right) \right] \right\}^{1/2}$$

- $N_{on}$ = # data in bin
- $N_{off}$ = # data outside bin
- $\alpha$ = ratio of $N_{on} / N_{off}$ for simulated isotropy
- $N_{bg}$ expectation

$N_{bg} = \alpha N_{data}^\text{data}$

- $N_{off}$ Normalized by exposure ratio

Test Used Previously by T.A. In:

INDICATIONS OF INTERMEDIATE-SCALE ANISOTROPY OF COSMIC RAYS WITH ENERGY GREATER THAN 57 EeV IN THE NORTHERN SKY MEASURED WITH THE SURFACE DETECTOR OF THE TELESCOPE ARRAY EXPERIMENT
**ESTIMATED BACKGROUND – EQUAL EXPOSURE**

- Equal exposure binning samples the sky equally.
  - “On” exposure such that bin size average = 15°, 20°, 25°, 30°
- Maximum Li-Ma significance for mean bin size of 25°

\[ \alpha = \frac{N_{on}^{MC}}{N_{off}^{MC}} = \text{constant} \]

\[ \langle \alpha \rangle = 0.0958 \quad \text{RMS } 1e - 04 \]

\[ \langle \text{Radius} \rangle = 25 \quad \text{RMS } 2 \]

\[ 25^\circ < \text{bin} > \]
RESULT
TWO ENERGY BIN LI-MA STATISTICS

Data: \( E \geq 57 \text{ EeV} \ (10^{19.75}) \) – a priori choice from previous studies

# Events in 25 <bin>

Pre-trial significance
TWO ENERGY BIN LI-MA STATISTICS

Data: $10^{19.1} \leq E \leq 10^{19.75}$

- Events in 25 <bin>
- Pre-trial significance

• Energy threshold scanned - $10^{19.0}, 10^{19.1}, 10^{19.2}, 10^{19.3}$ eV.
COMBINED LI-MA STATISTICS

High Energy HOTSPOT

- Maximum $\sigma_{local} = 7.11$ at 142° R.A., 40° Dec.
- $5^\circ$ from Energy Spectrum Anisotropy
- 16 degrees from supergalactic plane

Low Energy COLDSPOT

Cap with maximum joint significance (p-values multiplied)
HOT excess = 5.24 $\sigma_{local}$ and COLD deficit = -4.03 $\sigma_{local}$
EVIDENCE OF CAUSAL CONNECTION

Evidence for physical cause resulting in an excess at same point as deficit.

• Energy-Distance Correlation Anisotropy is direct evidence.

• Measured independently the Hotspot and Coldspot have the same size ~25°
SUPERGALACTIC PLANE SHIFT

Supergalactic magnetic sheet increase flux of post-GZK particles (E>50 EeV) and deflects (E<50 EeV) – suggested by (Biermann, Kang, Ryu)\(^1\)\(^2\)

Green line is linear in SG weighted by energy anisotropy \(\sigma^2\) of Hot/Coldspot points.

Result is SGP shifted -16°
TWO ENERGY BIN CORRELATIONS

These analyses can be done due to equal opening angle grid

Li-Ma $\sigma_{local}$ for Combined, Low Energy Bin, and High Energy bin

Correlated in time
(Integral day data $\sigma_{local}$ at max. point)

Correlated in Declination
($1^\circ$ Dec. bins average $\sigma_{local}$)

Correlated in Right Ascension
($1^\circ$ Dec. bins average $\sigma_{local}$)
TWO ENERGY BIN CORRELATIONS

Li-Ma $\sigma_{local}$ for Low Energy Bin, and High Energy bin

Excesses directly correlated with deficits (average of grid points within 0.1 $\sigma_{high}$ bins)

Linear Correlation $\sigma_{high} > 0 : -0.625$

# Grid points with Hot/Coldspot divided by # Hotspot
Versus high energy bin $\sigma_{high}$ cutoff
(100% of grid points with $\sigma_{high} > 3.24$ are a Hot/Coldspot)
EVENT DENSITY ASYMMETRY SIGNIFICANCE
EVENT DENSITY ASYMMETRY SIGNIFICANCE

- Testing MC trials for combined significance underestimates significance
  - Maxima with excess/deficit in both bins are not signatures of magnetic deflection
  - Significance of MC is found from separate energy bin $\sigma$ thresholds.

MC Sets scanned same bin sizes and energy thresholds as data

- 3 isotropic MC have equal or greater event density asymmetry out of 16 * 87.89 million
  - 5.4 $\sigma$ significance

MC sets outside of four bounds pass the test
Energy Spectrum Anisotropy significance: $3.74\sigma$ (parameters scanned and accounted for)

Energy-Distance Correlation Anisotropy significance: $3.37\sigma$ (parameters not scanned)

- **Stouffer’s Method combined significance**: $5.03\sigma$

Hot/Coldspot Anisotropy (a result of Spectrum AND Position Correlation): $5.4\sigma$
CONCLUSION

- Hot/Coldspot Event Density Asymmetry (energy-position correlation)
  - Post-trial $\sigma = 5.4$

- The previously reported Hotspot is correlated with a deficit of low energy events. This *observation* is suggestive of magnetic deflection.
PART TWO
COMPOSITION STUDY
PATTERN RECOGNITION ANALYSIS
&
QUALITY FACTOR ANALYSIS
• Protons deeper and wider than iron. Xmax (peak) gives composition information.

• Energy dependence of resolution is important if there is a change in composition.
PATTERN RECOGNITION ANALYSIS (PRA)

- No model needed to see increase and decrease in signal
- Fit shower profiles to triangles
  - Extract features from triangles. Describes shape of event. (Length of sides, angles, etc.)
- Brains are good at pattern recognition: use them to create training set of known good and bad events.
- Training set is used to find useful features, and cut values, for a yes/no determination.
  - The result agrees with the human observers on the 97.2% to 99.6% percent level.
Example of cut on two features extracted from triangles. (These two cut the most events)

- Obliqueness: perimeter/area of the large triangle.
- Right triangle area: $1/2 \times \text{Slantdepth} \times \text{Flux of the triangle sides}$.
BINARY PRA

- PRA determines whether an event has an acceptable profile and returns a binary yes/no answer.
FAILED PRA – PASSED GEOMETRY CUTS

Discrepancy is ~2 times the separation between means of Iron and proton primaries.

\[ E = 10^{18.3} \text{ eV}, \theta = 40.2 \text{ deg}, R_P = 17 \text{ m}, \text{SD/FD Core Diff} = 511, \]

Boundary Dist. = 2883 m, Tracklength = 13.4
Tight Geometry Cuts


Also used for:
A good start. How do you make it better? **MORE EVENTS**

- Maybe, we can lower our standards (or make the computer smarter than us) without compromising resolutions, resolution slopes, and biases.

- Instead of a yes/no answer a scale of event quality.
LOGISTIC REGRESSION

Finds weights, $\beta_j$, for prediction from features

$$\min_\beta J(\beta) = \sum_{j=1}^{N} [y_j \log p(\bar{x}_j) + (1 - y_j) \log(1 - p(\bar{x}_j))]$$

$$p_j(\bar{x}_j) = \frac{1}{1 + e^{-(\beta_0 + \beta \cdot \bar{x}_j)}}$$

$y_j$ (Binary PRA) (1 or 0) and $\bar{x}_j$ vector of triangle feature values for that event.
LOGISTIC REGRESSION

Logistic Function maps the range $(-\infty, \infty)$ to $[0, 1]$

Result is $p_j(t_j)$ the probability that the vector $\vec{x}_j$ comes from an event that is a ‘success’

$t_j = \beta_0 + \sum_{i}^{10} \beta_i \cdot x_{ji}$

Found weights $\beta_i$

$p_j(t_j) = \frac{1}{1 + e^{-(t_j)}}$

Logistic Function maps the range (-inf, inf) to [0,1]
EXAMPLE EVENT

Triangle Attributes

- Apex highest point = 1
- Bins before apex = 0.315
- Cubic term = 0.021
- Max. Sig. Diff. = 2.471
- Midsize length = 3.271
- Signal Mean = 3.524
- Norm. Missing = -1.721
- Apex angle/Hyp. = -1.665
- Left Oblique. = -3.292
- Large under right = -0.179

Fitted Weights

- $\beta_0 = -7.969$
- $\beta_1 = 3.474$
- $\beta_2 = 7.456$
- $\beta_3 = 42.286$
- $\beta_4 = 0.570$
- $\beta_5 = -0.054$
- $\beta_6 = 0.391$
- $\beta_7 = -0.242$
- $\beta_8 = 0.632$
- $\beta_9 = -3.351$
- $\beta_{10} = 2.068$

$t_j = \beta_0 + \sum_{i=1}^{10} \beta_i \cdot x_{ji} = 11.391$

$p_j(\bar{x}_j) = \frac{1}{1 + e^{-(11.391)}} = 0.99999$

- Highest energy event in data set.
- Has 5th highest quality factor at 0.99999
RESOLUTION CORRELATIONS WITH QUALITY FACTOR

RMS of difference between thrown and reconstructed values for proton MC

Xmax  Energy  Zenith Angle

*Very loose geometry cuts
QFA RESOLUTION IMPROVEMENT

Tight Geometry Cuts

Loose Geometry Cuts and PRA
Quality Factor describes how well events are seen by the FD

- Fairly linear correlation between Quality and RMS resolutions (and biases).
- Setting a QF threshold instead Binary PRA improves statistics.
7 YEAR GDAS(3-HOUR) DATA
4 YEAR QGSJETII-03
FLOATED X0 = -60, LAMBDA = 70

Quality Factor > 0.2 (for ~22 gm/cm² resolution), Energy > 18.4, Boundary Dist. > -500 m
SD/FD Core Difference < 1600 m, Zenith < 58, Geometry Fit Chi²/DOF < 5, and Xmax Bracketed.
DATA/MC COMPARISONS

1. **Proton** vs **Iron**
   - **Mean**: Proton: 18.8, RMS: 0.31
   - **Mean**: Iron: 18.8, RMS: 0.32

2. **Proton** vs **Iron**
   - **Mean**: Proton: 182, RMS: 108
   - **Mean**: Iron: 190, RMS: 108

3. **Proton** vs **Iron**
   - **Mean**: Proton: 92.8, RMS: 30
   - **Mean**: Iron: 93.8, RMS: 26

4. **Proton** vs **Iron**
   - **Mean**: Proton: 39.8, RMS: 12
   - **Mean**: Iron: 36.8, RMS: 13

---

**CVM p-value**
- Proton: 0.41
- Iron: 0.38

---

**CVM p-value**
- Proton: 0.23
- Iron: 0.54

---

**CVM p-value**
- Proton: 0.13
- Iron: 0.18

---

**CVM p-value**
- Proton: 5.2e-11
- Iron: 1.1e-07
XMAX DISTRIBUTIONS

<table>
<thead>
<tr>
<th>CVM p-value</th>
<th>Overall</th>
<th>18.4 to 18.6</th>
<th>18.6 to 18.8</th>
<th>18.8 to 19</th>
<th>19 to 19.2</th>
<th>Greater than 19.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proton: 0.0024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Iron: 3e-227</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Proton</td>
<td>Mean 755 RMS 58</td>
<td>Mean 748 RMS 63</td>
<td>Mean 755 RMS 66</td>
<td>Mean 750 RMS 51</td>
<td>Mean 759 RMS 61</td>
<td>Mean 759 RMS 61</td>
</tr>
<tr>
<td>Iron</td>
<td>Mean 686 RMS 34</td>
<td>Mean 674 RMS 33</td>
<td>Mean 674 RMS 33</td>
<td>Mean 682 RMS 31</td>
<td>Mean 682 RMS 31</td>
<td>Mean 682 RMS 31</td>
</tr>
<tr>
<td>CVM p-value</td>
<td>Proton: 0.01</td>
<td>Proton: 0.1</td>
<td>Proton: 0.26</td>
<td>Proton: 0.069</td>
<td>Proton: 0.059</td>
<td>Proton: 0.053</td>
</tr>
<tr>
<td>Proton</td>
<td>Mean 764 RMS 63</td>
<td>Mean 765 RMS 63</td>
<td>Mean 764 RMS 63</td>
<td>Mean 765 RMS 63</td>
<td>Mean 764 RMS 63</td>
<td>Mean 764 RMS 63</td>
</tr>
<tr>
<td>Iron</td>
<td>Mean 686 RMS 34</td>
<td>Mean 674 RMS 33</td>
<td>Mean 674 RMS 33</td>
<td>Mean 674 RMS 33</td>
<td>Mean 674 RMS 33</td>
<td>Mean 674 RMS 33</td>
</tr>
<tr>
<td>CVM p-value</td>
<td>Proton: 0.55</td>
<td>Proton: 0.059</td>
<td>Proton: 0.053</td>
<td>Proton: 0.059</td>
<td>Proton: 0.053</td>
<td>Proton: 0.053</td>
</tr>
<tr>
<td>Proton</td>
<td>Mean 763 RMS 62</td>
<td>Mean 761 RMS 52</td>
<td>Mean 772 RMS 45</td>
<td>Mean 774 RMS 59</td>
<td>Mean 774 RMS 59</td>
<td>Mean 774 RMS 59</td>
</tr>
<tr>
<td>Iron</td>
<td>Mean 686 RMS 63</td>
<td>Mean 690 RMS 30</td>
<td>Mean 690 RMS 30</td>
<td>Mean 690 RMS 30</td>
<td>Mean 690 RMS 30</td>
<td>Mean 690 RMS 30</td>
</tr>
</tbody>
</table>

Events: 613, 221, 135, 127, 70, 60
Xmax [gm/cm^2]: 800 to 1200
Number of Events: 200, 60, 50, 20, 10, 10
**MOMENTS**

**QSJETII-03**

Various models
COMPOSITION WITHOUT $<X_{\text{max}}>$
MOTIVATION

Results show model parameter uncertainty within a model results in $<X_{max}>$ uncertainty as large as difference between models.

Abbasi and Thompson

Variation between models: +/- ~15 g/cm$^2$

Data uncertainties:
- ~17 g/cm$^2$ systematic
- ~5 g/cm$^2$ statistical

Combined: ~30 g/cm$^2$ (23 without model var.)

Conclusion:
Uncertainties on $X_{max}$ distribution locations complicate the usual statistical inferences about composition

Dependence of $<X_{max}>$ on cross section, elasticity, and multiplicity at an energy of $10^{19.5}$ eV.
MOTIVATION

Data uncertainties:
• ~17 g/cm² systematic
• ~5 g/cm² statistical

Variation between models: +/- ~15 g/cm²

Combined: ~30 g/cm² (23 without model var.)

Conclusion:
Uncertainties on $X_{\text{max}}$ distribution locations complicate the usual statistical inferences about composition.
VARIANCE - NARROWING

Compare data to models

- Question: Significantly different variance?
  
  \[ H_0: \sigma_1^2 = \sigma_2^2 \]
  \[ H_a: \sigma_i^2 \neq \sigma_j^2 \]

- Method: O’Brien’s Test for Homogeneity of Variance

\[ W = \frac{(N - k) \sum_{i=1}^{k} N_i (\bar{Z}_i - \bar{Z})^2}{(k - 1) \sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2} \]

\[ Z_{ij} = \frac{N_i (N_i - 1.5)(y_{ij} - \bar{y}_i)^2 - 0.5\sigma_i^2(N_i - 1)}{(N_i - 1)(N_i - 2)} \]

P-value is calculated from \( F_{k-1, N-k} \), the \( F \) distribution with \( k-1 \) and \( N-k \) degrees of freedom.

- \( \text{RMS}(X_{\text{max}}) \) of data and QGSJETII-03
- Sampling issue or actual change?
VARIANCE - NARROWING

Compare data to models

- Significance of p-value that variance is the same
  - All models are in good agreement
  - No evidence for “narrowing” or change in composition
  - Statistically compatible with pure proton for any model at any energy

Compare data to data: test if $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$ for the 5 energy bins of data

Result: Significance of deviation is 0.97σ or 33% probability they are the same. Again, no statistical evidence for narrowed distribution
**MOMENTS \( \geq 2 \)**

- **Question:** Do two samples belong to the same location-family distribution?

  \( H_0: G(x) \) is sample CDF from \( F(z-a) \) & \( H(y) \) is sample CDF from \( F(z-b) \), for any \( a \) and \( b \)

  \[ x \in G(x) \sim F(z-a) \land y \in H(y) \sim F(z-b) \]

- **Method:** **L-test.** This is a more stringent test.

  \[
  L = \log \left\{ \min_{a \leq s \leq b} \frac{N_1 N_2}{(N_1 + N_2)^2} \sum_{k=1}^{N_1+N_2} [H(s)_k - G(s)_k]^2 \right\},
  \]

  \[
  \hat{F}(s)_k = \frac{1}{N_1} \sum_{j=1}^{N_1} I[(x_j - s) \leq z(s)_k], \quad \hat{G}(s)_k = \frac{1}{N_2} \sum_{j=1}^{N_2} I[y_j \leq z(s)_k], \quad z(s) = (x-s, y)
  \]

  Distribution of \( L \) is the Generalized Maximum Likelihood distribution
MOMENTS $\geq 2$

Tight geometry cuts and all energies

QGSJETII-03 distribution histograms and CDF’s shifted for best agreements
MOMENTS $\geq 2$

Compare data to models

- Significance of p-value that distributions are location family related
  - All models are in good agreement
  - No evidence for “narrowing” or change in composition
  - Statistically compatible with pure proton for any model at any energy
SHIFT PLOT

Shift plot using L-test by combining robust measure of bias and location family test

L-test shifts with O’Brien’s $\sigma$

L-test shifts with L-test $\sigma$ (stopped calculating at $6\sigma$)
• Statistical tests using distribution locations are inconclusive:
  • Stat. error, sys. error, model parameter uncertainty, model variation

• Higher moments agree between all models
  • *Data is statistically compatible with pure proton, at all energies, for all models*
  • Not compatible with iron.

• Significance of data being “narrowed” in RMS is $0.97\sigma$.

Next to do? Composition Anisotropy Using L-test?
THESIS CONCLUSIONS

• Hot/Coldspot Event Density Asymmetry Observed with 5.4σ significance
  • Energy Spectrum Anisotropy with 3.74σ
  • Energy-Distance Correlation Anisotropy with 3.4σ
  • Suggests magnetic deflection of source by possible supergalactic fields

• Higher moments of Xmax distributions agree between all models
  • Data is statistically compatible with pure proton, at all energies, for all models
  • Not compatible with iron.

• Significance of data being “narrowed” in RMS is 0.97σ.

\[ E < \frac{E_{\text{max}}}{10^{18} \text{ eV}} \cong \frac{1}{2} \beta \cdot Z \cdot \frac{B}{\mu G} \cdot \frac{L}{\text{kpc}} \]

\[ \delta = \frac{S}{R_{\text{Larmor}}} = 0.5^\circ Z \frac{L}{\text{kpc}} \frac{B}{\mu G} \frac{10^{20}}{E} \]

This information should be useful for informing future models of magnetic fields and sources

*Kernel Density Estimation Hotspot: 3.65σ
PART ONE
ADDITIONAL MATERIAL
CORSIKA Simulated Air Shower \(10^{15} \text{ eV} \) 45° inclination

Red – e\(^{+/-}\), γ
Green – μ\(^{+/-}\)
Blue – Hadrons (π\(^{0/+/-}\), K\(^{0/+/-}\), p, n)

Gaisser-Hillas Parameterization

\[
N(X) = N_{\text{max}} \left( \frac{X - X_0}{X_{\text{max}} - X_0} \right)^{X_{\text{max}} - X_0} \exp \left( \frac{X_{\text{max}} - X}{\lambda} \right)
\]

- Match fit to real event data.
Largely from coulomb multiple scattering of electrons

Nishimura-Kamata-Geisen (NKG) formula.
Lateral density of electrons as function of shower age

\[
\rho(r) = \frac{N}{r^2} f\left(s, \frac{r}{r_M}\right)
\]

\[
f\left(s, \frac{r}{r_M}\right) = \left(\frac{r}{r_M}\right)^{s-2} \left(1 + \frac{r}{r_M}\right)^{s-4.5} \frac{\Gamma(4.5 - s)}{2\pi \Gamma(s) \Gamma(4.5 - 2s)}
\]

\[s = \frac{3X}{2X + X_{max}}\]

\[R_M = \frac{21 \text{ MeV}}{E_c} \lambda \approx 9 \text{ gm/cm}^2\]

Shower Age: 1 is Xmax

Moliere radius
RECONSTRUCTION

Chi-Square Minimization of Parameters: $\psi, \theta, R_P, R_x, R_y$

$$\chi^2 = \chi^2_{MD\text{Timing}} + \chi^2_{SD\text{Timing}} + \chi^2_{SD\text{Core}}$$

First find SDP:

SDP Normal  Tube Direction

SD Virtual Tube Timing:

Timing to Minimize:

$$t_{SD} = t_{SD\text{Trig}} + \frac{SD_{Dist}}{c}$$

$$t_i = T_{Rp} + \frac{R_P}{c} tan \left( \frac{\pi - \psi - \chi_i}{2} \right)$$

$$\chi^2_{Core} = \sum_{i=1}^{2} \frac{||R_i - R_{COG}||^2}{\sigma^2_{R_{COG}}}$$
ENERGY SPECTRUM ANISOTROPY
ADDITIONAL MATERIAL
MC DISTRIBUTION OF HITS

Shows small amount of declination bias in the analysis
MC DISTRIBUTION OF HITS
MC CHI^2 DISTRIBUTION AT DATA MAX SIGMA POINT

138.8 R.A. 44.8 Dec.
19.2 energy cut
30 deg binning

"Chi square" distribution of MC sets

Data chi square: 78.3 for 14 energy bins

- There are two additional degrees of freedom:
  - Background Fluctuation
  - Rebinning
MC DISTRIBUTIONS AT DATA MAX SIGMA POINT

138.8 R.A. 44.8 Dec.
19.2 energy cut
30 deg binning

MC N inside is Poisson: 163.8 +/- 12.0
(sqrt(163.8) = 12.8)

MC N_bg background is not Poisson: 163.8 +/- 1.7
Fluctuation is sqrt(N)*0.14 exposure ratio exactly

This is the same background fluctuation Li-Ma uses
MC CHI^2 DISTRIBUTION AT DATA MAX SIGMA POINT

138.8 R.A. 44.8 Dec.
19.2 energy cut
30 deg binning

“Chi square” distribution of MC sets
with no background fluctuation or rebinning

547 MC have infinite chi^2 due to no rebinning

MC sets with 14 energy bins
Closest to chi^2 with 14 degrees of freedom
Could systematics cause events to migrate from Coldspot to Hotspot?

Energy is reconstructed by Zenith angle and s800 signal
- Zenith agrees very well. Systematic must come from s800
- s800 would have to be increased by 139% for hotspot to be systematic from the coldspot

E \geq 57 \text{ EeV events: } \sim 14 \text{ over } N_{bg} \text{ or } 3.6N_{bg}
20 \leq E < 57 \text{ EeV events: } \sim 21 \text{ under } N_{bg} \text{ or } 0.57N_{bg}
OTHER SYSTEMATIC CHECKS

- Seasonal and hourly energy corrections result in little change to joint significance.
- Anti-Sidereal time results in no significant excesses, deficits or combinations.

\[
20 \leq E < 57 \text{ EeV Anti-Sidereal}
\]

\[
E \geq 57 \text{ EeV Anti-Sidereal}
\]
ENERGY-DISTANCE CORRELATION
ADDITIONAL MATERIAL
CHECK FOR GOOD BEHAVIOR

- Correlation coefficients follow the $t$ Location-Scale distribution
  - Literature states correlations of correlations should have a wider distribution

- $p$-values should be Uniform (Null – single correlation)
  or a Beta distribution (prior information – second correlation)

Ranked correlation $\tau$
Fits well

Double Ranked Correlations $\tau | \tau$
 Doesn’t behave properly
Not Used

$\rho | \tau, p$
Fits distribution well
Used for significance test

Histograms of 852 test points from 300 MC – 2.6e5
CHECK FOR GOOD BEHAVIOR

- Correlation coefficients follow the t Location-Scale distribution
  - Literature states correlations of correlations should have a wider distribution

- p-values should be Uniform (Null – single correlation)
  or a Beta distribution (prior information – second correlation)

Ranked correlation p-Value Uniform Distribution

Double Ranked Correlations p-Value Beta Distribution Not Used

\[ \rho_{\tau|p}'s \text{ p-Value} \]

Fits Beta Distribution Better Used for significance test

Histograms of 852 test points from 300 MC – 2.6e5
INTEGRAL DAY SIGNIFICANCE

- Location of maximum colored/sized by MJD
- General location is found within ~3 years

8 year estimate ~4\sigma_{global}

- Blue line is linear fit — 0 to 7 years
- Red line is linear fit — 5th to 7 years

- Maximum \( \sigma_{local} \) on map
- Linear correlation 0.905 (0.935 after 5th year)
YEAR BY YEAR TREND

8 year estimate $\sim 4\sigma_{\text{global}}$

- Median sliding 1 year change
  - 0 to 7 years $\sim +0.7 \sigma$/year
  - 5 to 7 years $\sim +1.8 \sigma$/year

- 2,107 sliding 1 year $\sigma$ differences (0 to 7 years)
  - 1,554 $\sigma$ increases
  - 553 $\sigma$ decreases

We only use integral year data.
DECLINATION COUNT DISTRIBUTION

This is for **negative** $\rho_{|r|,\rho}$ only. The data was negative.

Position is dependent on over/under-density but significance is not.
Position is also more sensitive to energy anisotropy as shown by integral day data figures.
This is for negative $\rho_{|\tau|,\rho}$ only. The data was negative. The data was negative.

Position is dependent on over/under-density but significance is not.

Position is also more sensitive to energy anisotropy as shown by integral day data figures.
SEASONAL CORRECTION TEST
Each test point \((i)\) calculate Kendall’s \(\tau\) correlation \(F_i[E_j(E_j > E_i), \theta_{ij}(E_j > E_i)]\)

Choosing events with Energy > test point energy - removes adjacent double counting.

- **Negative Correlation:** Energy Increases → Angle decreases
- **Positive Correlation:** Energy Increases → Angle Increases

- Size proportional to \(1/p\)-Value
- Color is Opening angle/Energy correlation

**DATA RESULT**

- **Negative Correlation:** Energy Increases → Angle decreases
- **Positive Correlation:** Energy Increases → Angle Increases

Each Test Point sample size is different

Data: 833 events
E >20 EeV after adjustment

4 MOST SIGNIFICANT POINTS

Energy Vs Distance

\[ E \geq E_i \]

Test Point location and Energy Cut

\[
\begin{array}{cc}
\text{tau} & \text{pval} \\
-0.198 & 2.80E-05 \\
-0.187 & 5.43E-05 \\
-0.119 & 3.78E-04 \\
-0.163 & 4.49E-04 \\
\end{array}
\]
RANKING

• An ordered list of magnitude.
• Ranking removes functional form of dependence.
• Lowest variable value = 1.
• Highest variable value = N events.

Example: Energy vs log(energy)

Linear Correlation: 0.92

Rank Correlation: 1
LINEAR CORRELATION (PEARSONS)

\[
\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}
\]

\[
r = r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

Measurement of linear dependence
RANK CORRELATION

- All values are ranked. Kendall’s correlation is used.
  \[
  \tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2} n(n - 1)}
  \]

- Ranked correlations are permissive: any perfectly monotonic function \( F(x,y) \) results in correlation = 1. **Removes model assumption**

  - **p-Values**
    - calculated by permutation
    - null hypothesis is zero correlation.
    - p-Value is probability correlation is zero.

Source: wikipedia

- Linear Corr: 0.903
  Outlier decreases corr: ~0.02

- Rank Corr: 0.9994
  Outlier decreases corr: 0.0006

Robust against outliers
GLOBAL POINT SIGNIFICANCE

- 1,000,000 Isotropic MC maps the size of data
- Count maps with at least one test point with:
  - $|\tau| \geq |\tau_{data}|$ \& \(\text{sign}(\tau) = \text{sign}(\tau_{data})\) \& \(p \leq p_{data}\)
  (At least as ordered change in same direction, with greater or equal samples)

- Size proportional to $1/p_{global}$
- Color is Global Significance
- Highest Significant test point: $2.7\sigma$

Individual correlations clearly don't tell the whole story...
EXAMPLE MC MAPS

Distribution of p-vals and tau least like data

Distribution of p-vals most like data

Distribution of tau most like data

Most points like best p-val point from data. (2 points)
DATA CUMULATIVE TIME QUANTILES

Quartile 1

Quartiles 1 and 2

Quartiles 1 to 3

All
DIFFERENT DATA SUBSETS

7 Year Kawata-san data with additional cuts

7 Year Dmitri’s data (tighter cuts)

5 Year Kawata-san hotspot paper data

7 Year Kawata-san ICRC data
HOT/COLDSPOOT SUMMARY ANALYSIS
ADDITIONAL MATERIAL
Inside hotspot there is possibly something different with zenith, energy, and RA.

One random isotropic MC map the size of the data.
COMPARISON WITHIN HOT/COLD SPOT – E > 57 VS E < 57
S800 AND CLUSTERS – SHOULD NOT BE THE SAME
COMPARE COLDSPOT TO ISOTROPIC MC

Energy and RA are a bit different.
CVM p-value
0.62

Hotspot

Coldspot

Hotspot
Mean 5.59e+04
RMS 9.4e+02

Coldspot
Mean 5.58e+04
RMS 8.2e+02
COMPARE HOTSPOT TO ISOTROPIC MC
COMPOSITION
ADDITIONAL MATERIAL
Examples of attributes

- At least 2 bins before apex and either end.
- Cubic term of quadratic fit used to find triangle apex.
- Size of small side of large triangle.
- Standard deviation of signal flux.
- Normalized maximum missing slant depth in profile.
- Obliqueness (perimeter/area) of large triangle.
- Allowed missing area between bins.
- Normalized Largest side of under right triangle.
- Ratio of normalized largest side of large triangle to apex angle.
RECONSTRUCTED VS. THROWN

- (Reconstructed – Thrown) Vs. Quality
- Minimum cuts applied to make limits of MC and data the same:
  - \( \log_{10}(\text{energy}) > 18.2 \) & \( \text{boundarydist} > -1500 \) & \( \text{corediff} < 2500 \) & \( \text{zenith} < 60 \)

![Graphs showing RMS vs. Quality for Rp and Zenith](image)
RECONSTRUCTED VS. THROWN

- (Reconstructed – Thrown) Vs. Quality
- Minimum cuts applied to make limits of MC and data the same:
  - \( \log_{10}(\text{energy}) > 18.2 \) & \( \text{boundarydist} > 1500 \) & \( \text{corediff} < 2500 \) & \( \text{zenith} < 60 \)

Less spread at higher quality.

Energy

Xmax
RESOLUTION VS. ENERGY VS. Q THRESHOLD

- For $Q > 0$
- For $Q \geq 0.1$
RESOLUTION VS. ENERGY VS. Q THRESHOLD

Q>=0.4

Q>=0.5
RESOLUTION VS. ENERGY VS. Q THRESHOLD

Q>=0.6

Q>=0.7
Resolution with respect to energy flattens with increasing Quality

Q$\geq$0.8
SYSTEMATIC ERRORS

\[ \langle X_{\text{max}} \rangle = 751 \pm 16.3 \text{ sys. } \pm 9.4 \text{ stat. } \text{gm/cm}^2 \text{ at } \log_{10}(E) = 19 \]

\( \langle X_{\text{max}} \rangle \) Systematic errors include:
- Mirror alignment (known to \( \pm 0.05^\circ \)): \( \pm 2.6 \text{ gm/cm}^2 \)
- Atmosphere Density (US 1976 Standard Vs. Yearly Ave. Radiosonde): \( \pm 11.7 \text{ gm/cm}^2 \)
- Vertical Aerosol Optical Depth (VAOD) Nightly Variation: \( \pm 2 \text{ gm/cm}^2 \)

Figure 1. Differences of atmospheric depth from the US-SA model (left: average, right: standard deviation)

Systematic effects of Cerenkov light subtraction is negligible due allowed hybrid shower directions.
**ROBUST MEASURE OF BIAS**

L-test shift - robust bias measure for skewed distributions  
(Distance between population modes/locations)

**Procedure:** 5000 random number sets from fitted distributions and measure distances

**Test:** Measure L-test Shifts, Difference of sample modes, Difference of medians, and Difference of Means

---

**EPOS PROTON**

18.8 to 19

Mode distance of GEV fit distributions

26.30 +/- 4.68 g/cm\(^2\) (fit uncertainty)

---

**Data Set**

#data = 127, #MC = 937

- Shift: 25.2 g/cm\(^2\)
- Mode: 95.5 g/cm\(^2\)
- Median: 28.2 g/cm\(^2\)
- Mean: 22.2 g/cm\(^2\)

**Distances of 5000 MC**

- Shift: 26.7 RMS 4.8 g/cm\(^2\)
- Mode: 5.9 RMS 12.3 g/cm\(^2\)
- Median: 27.3 RMS 6.1 g/cm\(^2\)
- Mean: 27.9 RMS 5.6 g/cm\(^2\)
**ROBUST MEASURE OF BIAS**

L-test shift - robust bias measure for skewed distributions
(Distance between population modes /locations)

**EPOS PROTON**

19 to 19.2

Mode distance of GEV fit distributions
34.78 +/- 5.68 g/cm^2

Procedure: 5000 random number sets from fitted distributions and measure distances

Distances of 5000 MC

- Shift: 36.1 RMS 5.8 g/cm^2
- Mode: 16.4 RMS 12.7 g/cm^2
- Median: 36.4 RMS 7.3 g/cm^2
- Mean: 39.8 RMS 6.4 g/cm^2

Data Set

#data = 70, #MC = 549

- Shift: 34.83 g/cm^2
- Mode: 114.87 g/cm^2
- Median: 30.45 g/cm^2
- Mean: 39.77 g/cm^2

**ROBUST MEASURE OF BIAS**

L-test shift - robust bias measure for skewed distributions
(Distance between population modes /locations)
ROBUST MEASURE OF BIAS

L-test shift - robust bias measure for skewed distributions
(Activation between population modes /locations)

**EPOS IRON**
18.8 to 19

Mode distance of GEV fit distributions
-37.4 +/- 4.4 g/cm^2

Data Set
- #data = 127, #MC = 1100
- Shift: -46.9 g/cm^2
- Mode: +34.5 g/cm^2
- Median: -48.0 g/cm^2
- Mean: -52.5 g/cm^2

Distances of 5000 MC
- Shift: -43.1 RMS 5.2 g/cm^2
- Mode: -28.6 RMS 11.6 g/cm^2
- Median: -44.3 RMS 5.8 g/cm^2
- Mean: -52.3 RMS 5.2 g/cm^2
ROBUST MEASURE OF BIAS

L-test shift - robust bias measure for skewed distributions
(Distance between population modes /locations)

EPOS IRON
19 to 19.2

Mode distance of GEV fit distributions
-28.92 +/- 5.39 g/cm^2

Mode distance of GEV fit distributions
-28.92 +/- 5.39 g/cm^2

#data = 70, #MC = 685
- Shift: -37.3 g/cm^2
- Mode: +16.8 g/cm^2
- Median: -39.9 g/cm^2
- Mean: -40.7 g/cm^2

Distances of 5000 MC
- Shift: -33.87 RMS 6.16 g/cm^2
- Mode: -29.3 RMS 12.6 g/cm^2
- Median: -34.26 RMS 6.86 g/cm^2
- Mean: -40.22 RMS 5.98 g/cm^2
KERNEL DENSITY ESTIMATION
ADDITIONAL MATERIAL
KERNEL DENSITY ESTIMATION

Test statistic: Wald’s Proportion test

\[ Z = \frac{\hat{p} - p_{bg}}{\sqrt{\hat{p}(1-\hat{p})}} \]  

Flattest Dec. Response

Optimal von-Mises-Fisher kernel concentration for PDF found automatically for data and MC

5 year tight cuts shown

Post-trial sigmas
5 to 9 year
Loose cuts: 3.89, 4.36, 3.84, 2.92, 2.78
Tight cuts: 3.72, 4.39, 3.81, 3.06, 2.98
KDE DECLINATION RESPONSE

Next best statistic: $p/sqrt(p_{bg})$
Equal opening angle grid. \( p_{bg} \) calculated with trigger times for each year.

KDE PDF TIGHT CUTS

5 year

6 year

7 year

8 year

9 year
Equal opening angle grid. $p_{bg}$ calculated with trigger times for each year.

**KDE PDF LOOSE CUTS**
EVEN MORE HOT/COLD ADDITIONAL MATERIAL
HOT/COLD SPOT
-SUMMER/WINTER AND NIGHT/DAY

Jon Paul Lundquist
Outside hot/cold spot

Inside hot/cold spot

Energy distributions agree within statistics
FIELD OF VIEW PROBLEM

Inside hot/cold spot number of Events per hour is different than overall sky due to TA decl. = 40

Per hour per month 24*12 = 288 frames
FIELD OF VIEW PROBLEM

Night – January to December
Uneven in RA

Day – July to June (6 months offset)
Uneven in Declination
SUMMER/WINTER ENERGY DISTRIBUTION COMPARISON

Energy distributions agree within statistics.
SEASONAL ENERGY CORRECTION

- Energy correction found from reconstructed MC using Elko radiosonde data (D.Ivanov)
- Lateral dist. change from atmos. temperature changes

Correction +/- 7%

Affects 20 EeV cut and 57 EeV energy split

Energy Vs Month

# Events per hour by Month flattens out
SUMMER/WINTER ENERGY AFTER CORRECTION

Outside hot/cold spot
Agreement improved

Energy distributions agree within statistics

Inside hot/cold spot
Agreement improved
DAY/NIGHT ENERGY AFTER CORRECTION

Outside hot/cold spot

Inside hot/cold spot

Agreement improved

Energy distributions agree within statistics
852 events

Original $\sigma$ Combined
Max: 5.92

At max:
R.A. = 139
dcl. = 48
Cold = -3.25
Hot = 4.54

833 events

Seasonal Energy Corrected $\sigma$ Combined
Max: 5.36

At max:
R.A. = 137
dcl. = 48
Cold = -2.80
Hot = 4.16

5000 random samples of 833 events:
Combined Median: 5.90 - 0.14 + 0.12
Corrected Combined is 3.86 error bars from median.
SEASONAL ENERGY CORRECTION

852 events
Original $\sigma E < 57$ EeV
Max: 4.61

833 events
Seasonal Corrected $\sigma E < 57$ EeV
Max: 4.22

852 events
Original $\sigma E > 57$ EeV

833 events
Seasonal Corrected $\sigma E > 57$ EeV
HOURLY ENERGY CORRECTION

- Energy correction found from hourly events rates (D. Ivanov)
- Lateral dist. change from atmos. temperature changes

Correction +/- 5% after seasonal correction

Affects 20 EeV cut and 57 EeV energy split

Energy Vs Hour
Outside hot/cold spot
Agreement improved

Inside hot/cold spot
Agreement improved

Energy distributions agree within statistics
DAY/NIGHT ENERGY AFTER HOURLY CORRECTION

Outside hot/cold spot

Inside hot/cold spot

Agreement improved

Energy distributions agree within statistics
HOURLY ENERGY CORRECTION

852 events

Original $\sigma$ Combined
Max: 5.92

At max:
R.A. = 139
decl. = 48
Cold = -3.25
Hot = 4.54

844 events

Hourly Energy Corrected $\sigma$ Combined
Max: 5.86

At max:
R.A. = 139
decl. = 48
Cold = -3.25
Hot = 4.48

5000 random samples of 844 events:
Combined Median: 5.90 - 0.07 + 0
Corrected Combined is 1.7 error bars from median.
**HOURLY ENERGY CORRECTION**

- **COLDSPOT**
  - Original $\sigma E < 57 \text{ EeV}$
  - Max: 4.61
  - 852 events

- **HOTSPOT**
  - Original $\sigma E > 57 \text{ EeV}$
  - Max: 4.52
  - 844 events

- **Hourly Corrected $\sigma E < 57 \text{ EeV}$**
  - 844 events

- **Hourly Corrected $\sigma E > 57 \text{ EeV}$**
  - 852 events
CONCLUSION

- Energy distributions between day/night and summer/winter agree within statistics.
- After MC derived energy correction energy distributions still agree.
- Hot/Coldspot is stable after energy correction. Affects both hotspot and coldspot almost equally.
APPENDIX
SPLIT CONCLUSION

No statistically significant difference from full data set is found by splitting data in half.
SPLIT SUMMER – WINTER

368 events
Winter σ Combined
Max: 5.52

At max:
R.A. = 139
decl. = 48
Cold = -3.90
Hot = 3.89

Rand data (same #) σ, Combined
Hot/Cold Median: 4.64 – 0.55 + 0.63
Significance higher than random sampling by 1.4 sigma.

484 events
Summer σ Combined
Max: 4.08

At max:
R.A. = 146
decl. = 42
Cold = -1.78
Hot = 3.24

Random data (same # events) σ, Combined
Hot/Cold Median: 5.01 – 0.56 + 0.55
Significance lower than random sampling by 1.7 sigma.
SPLIT SUMMER – WINTER

Winter $\sigma E < 57$ EeV

- 368 events

Winter $\sigma E > 57$ EeV

- Max. Hot: 4.43

Summer $\sigma E < 57$ EeV

- 494 events

Summer $\sigma E > 57$ EeV

- Max. Hot: 3.24
Summer – April, 15th to October, 15th.

5000 random selections of 494 events

HOT/COLD SOURCE SIGNIFICANCE – SUMMER ONLY

437 events
σ E < 57 EeV
Min Cold: -3.46

47 events
σ E > 57 EeV
Max Hot: 3.24

At Hot/Cold Spot: -1.99

At Hot/Cold Spot: 1.95
HOT/COLD SOURCE SIGNIFICANCE – SUMMER ONLY

- Summer – April, 15th to October, 15th.

494 events
Summer $\sigma$ Combined
Max Hot/Cold: 4.08

Random data (same # events) $\sigma$, Combined
Hot/Cold Median: 5.01 – 0.56 + 0.55

Significance lower than random sampling by 1.7 sigma.
Change not significantly different from random sampling
HOT/COLD SOURCE SIGNIFICANCE – WINTER ONLY

- Winter – October 16\textsuperscript{th} to April, 14th

5000 random selections of 368 events

331 events
\[ \sigma E < 57 \text{ EeV} \]
Min Cold: -3.3

37 events
\[ \sigma E > 57 \text{ EeV} \]
Max. Hot: 4.43

At Hot/Cold Spot -2.69

At Hot/Cold Spot 4.42
Winter – October 16\textsuperscript{th} to April, 14th

Rand data (same #) $\sigma$, Combined
Max: $4.64 - 0.55 + 0.63$

Significance higher than random sampling by 1.4 sigma. Change not significantly different from random sampling.
DAY-NIGHT SPLIT
SPLIT DAY – NIGHT

432 events
Day σ Combined
Max: 3.95

At max:
R.A. = 146
dec. = 42
Cold = -3.30
Hot = 3.29

Rand data (same #) σ, Combined
Hot/Cold Median: 4.82 - 0.56 + 0.60
Significance lower than random sampling by 1.6 sigma.

420 events
Night σ Combined
Max: 5.82

At max:
R.A. = 139
dec. = 48
Cold = -1.42
Hot = 3.29

Rand data (same #) σ, Combined
Hot/Cold Median: 4.87 - 0.57 + 0.58
Significance higher than random sampling by 1.7 sigma.
SPLIT DAY – NIGHT

432 events
Day $\sigma E < 57 \text{ EeV}$

Night $\sigma E < 57 \text{ EeV}$

Coldspot

Day $\sigma E > 57 \text{ EeV}$
Max: 3.63

Night $\sigma E > 57 \text{ EeV}$
Max: 4.37

Hotspot
HOT/COLD SOURCE SIGNIFICANCE – DAY ONLY

- Day – 9am to 9pm
  - 15 to 3 GMT

386 events
Day $\sigma E < 57$ EeV

46 events
Day $\sigma E > 57$ EeV

Coldspot
At Hot/Cold Spot: -1.39

Hotspot
At Hot/Cold Spot: 2.00
HOT/COLD SOURCE SIGNIFICANCE – DAY ONLY

- Day – 9am to 9pm

432 events
Day σ Combined
Max: 3.95

Rand data (same #) σ, Combined
Max: 4.87 - 0.57 + 0.58

Significance lower than random sampling by 1.6 sigma.
Change not significantly different from random sampling
HOT/COLD SOURCE SIGNIFICANCE – NIGHT ONLY

- Night – 9pm to 9am
  - At Hot/Cold Spot: -3.30

382 events
Night $\sigma E < 57$ EeV

38 events
Night $\sigma E > 57$ EeV

Coldspot
Hotspot

At Hot/Cold Spot: 4.37
HOT/COLD SOURCE SIGNIFICANCE – NIGHT ONLY

- Night – 9pm to 9am

420 events
Night $\sigma$ Combined
Max: 5.82

Rand data (same #) $\sigma$, Combined
Max: 4.82 – 0.56 + 0.60

Significance higher than random sampling by 1.7 sigma. Change not significantly different from random sampling.
3 COMPONENT COMPOSITION – FIT XMAX AND S800

Jon Paul Lundquist
METHOD

- Use Proton, Iron, and Nitrogen primaries.
- Find best fit to Xmax and s800 distributions simultaneously
  - Maximize combined p-Value (p1*p2) from CVM test.
- For each energy bin
  - For each ratio
    - Calculate combined p-Value for 100 different samples Ndata*5.
    - Find ratio which maximizes the mean combined p-value.
    - Iterate 100 times (using a different data sample with replacement – bootstrap) to find error on ratio which maximizes the p-value.
RESULT

Consistent with zero iron. Consistent with 2 component fit using only $X_{\text{max}}$
PVALUES

Distribution p-Values Vs Energy

Log10(E/eV) > 19.2 – 1 iteration

Xmax p-Values

s800 p-Values