# Staying Sane in an Insane World: Novel Analyses for Cosmic Ray Experiments

John A.J. Matthews

johnm@phys.unm.edu

University of New Mexico

Albuquerque, NM 87131

**FOR**

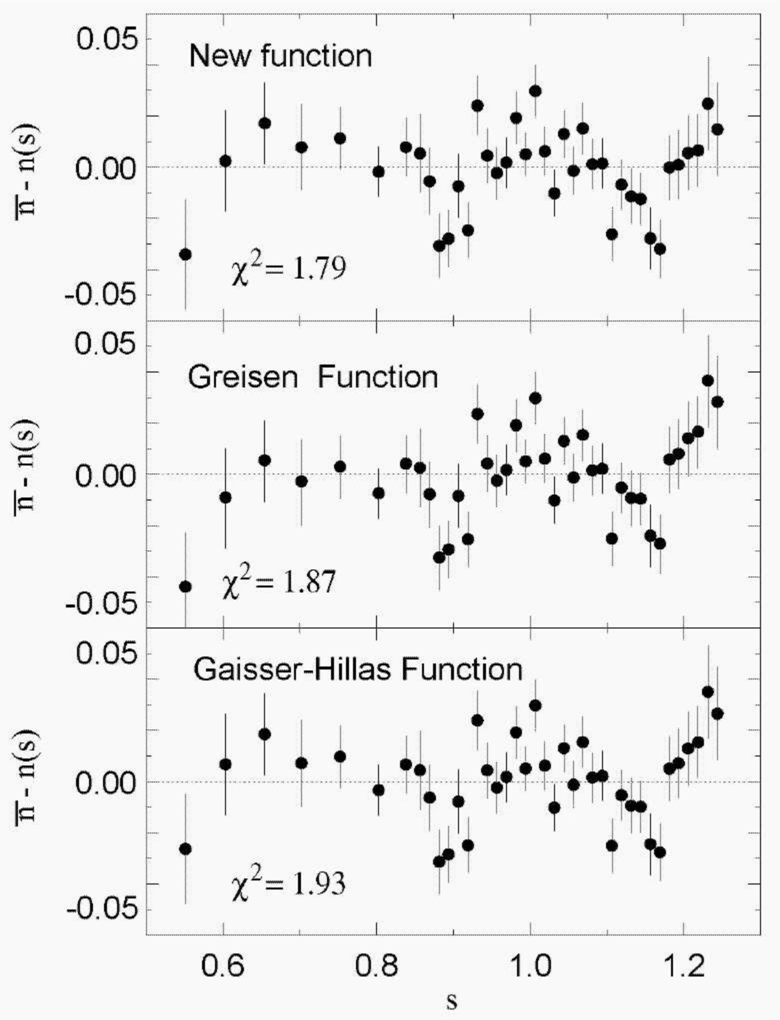
Bernie Becker, Michael Gold, Doug Hague, and Bob Mesler

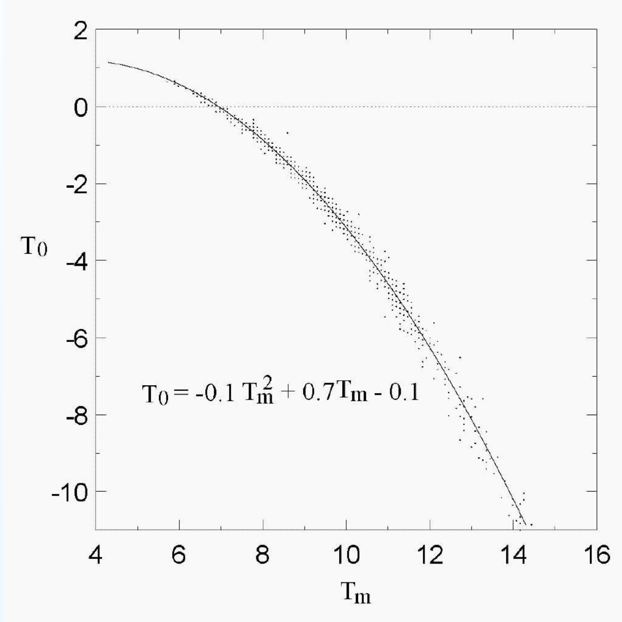
# (I) A *new* Parametrization of CR Shower Profiles

- Recall that CR air showers, detected via the air fluorescence technique, are reconstructed using functions that parametrize the longitudinal profile of each shower.
- The Auger reconstruction uses the Gaisser-Hillas 4-parameter form.
- The HiRes group has used both the Gaisser-Hillas form and a 3-parameter Gaussian-in-Age form.
- Historically analytic shower theory suggested yet other forms; the best known is a 3-parameter form popularized by Greisen.
- Is there an *optimal* form? How different are the familiar forms?
- Let this *percolate* for about 8 years ..

.

# HiRes-Prototype (2001) Study of Shower Profiles

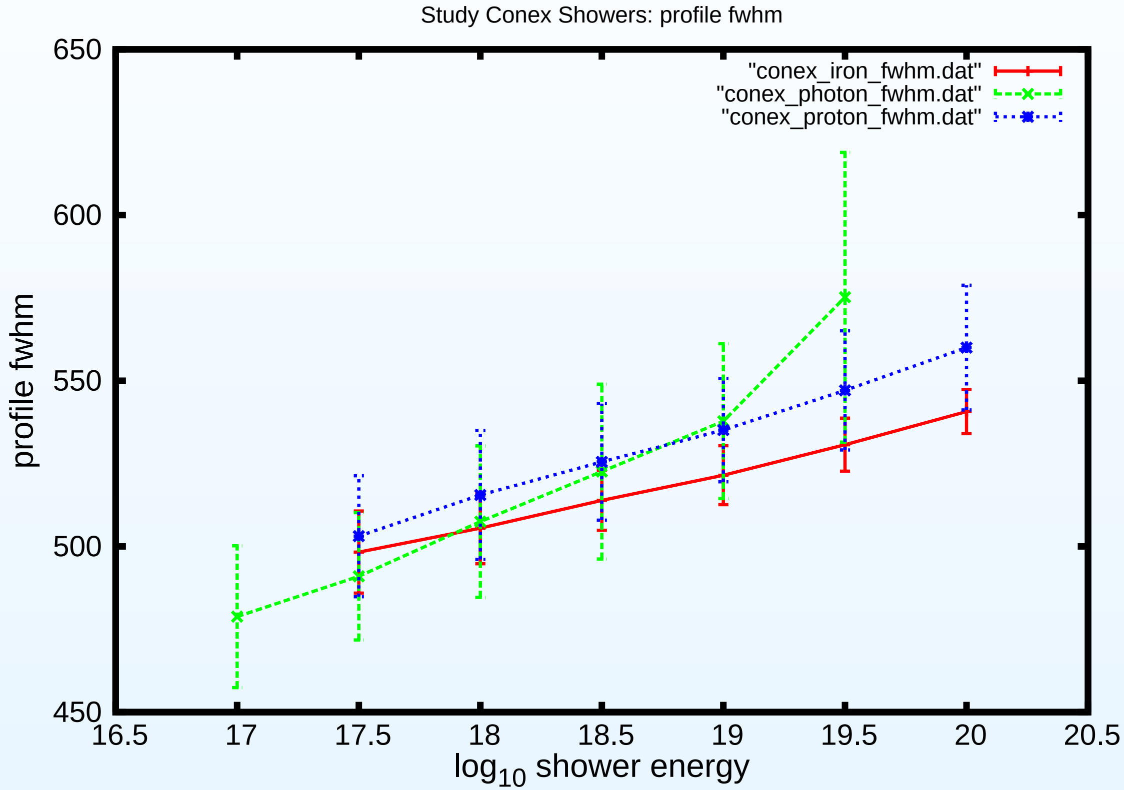



- Left plot: Residuals from comparison of HiRes-Prototype composite shower profile to GIA, Greisen and GH parametrizations. Events have $10^{17} \leq E \leq 10^{18}$eV.

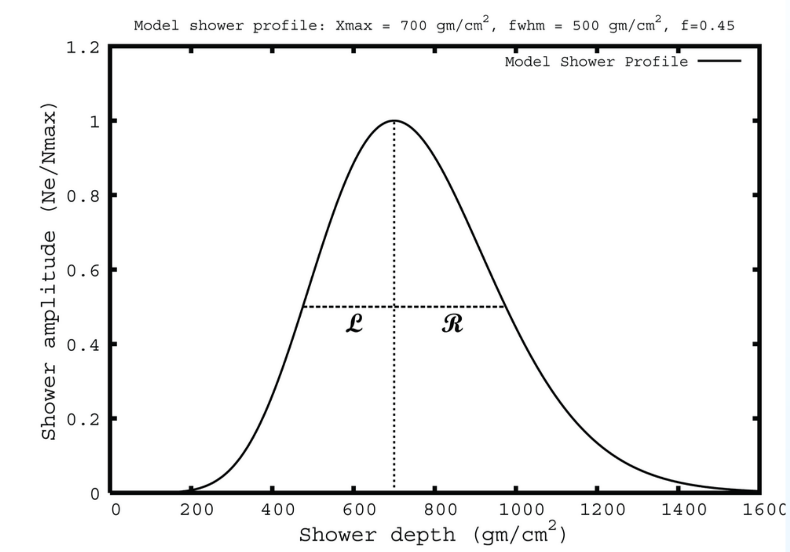
- Top plot: Observed correlation between GH parameters: $T_0 = \frac{X_0}{\lambda}$ and $T_m = \frac{X_{max}}{\lambda}$

# HiRes (2001) Corsika Study of Shower Profiles




- Conclusion: GH and GIA functions described Corsika showers comparably well
- However GIA was preferable as it required only 3 parameters
- The Monte Carlo study observed the near equality of the width at half-maximum, $fwhm$, of proton, iron and photon showers but did not exploit this fact

# A New Approach to Shower Profiles (I)




- How different are the different: GH, GIA and Greisen profiles?
- Are 3-parameters indeed sufficient or are 4-parameters needed?
- Can we profit from similarity of shower $fwhm$?
  1. reformulate GH, GIA and Greisen profiles based on: $N_{max}$, $X_{max}$, $fwhm \equiv \mathcal{L} + \mathcal{R}$ and shower asymmetry $f \equiv \mathcal{L}/(\mathcal{L} + \mathcal{R})$.
  2. then all profiles depend on two dimensionless ratios: $\epsilon \equiv \frac{\Delta}{W}$ where $\Delta = X - X_{max}$, $W \equiv X_{max} - X_0 = \frac{fwhm}{R(f)}$, and $\xi \equiv \frac{W}{\lambda}$, $\sigma$ or $\frac{W}{p_{36.7}}$ where $\xi$ depends only on the asymmetry $f$.

# A New Approach to Shower Profiles (Ia)

- Gaisser-Hillas:

$$N(X)_{GH} \;=\; N_{max}\,\Big(\frac{X - X_0}{X_{max} - X_0}\Big)^{\frac{X_{max}-X_0}{\lambda}}\; e^{-\frac{X - X_{max}}{\lambda}}$$

with parameters, $N_{max}$, $X_{max}$, $X_0$ and $\lambda$. Note: $X_0$ is <u>not</u> the actual *start* of the shower!

- Gaussian-in-Age:

$$N(X)_{GIA} \;=\; N_{max}\; e^{-\frac{1}{2}\,(\frac{s-1}{\sigma})^2}$$

with (nominal) parameters, $N_{max}$, $X_{max}$ (as $s = \frac{3X}{X+X_{max}}$) and $\sigma$. Note: this assumes the shower starts at $X = 0$.

- Greisen:

$$N(t)_{Greisen} \;=\; N_{max}\; e^{(\;t(1-\frac{3}{2}ln(s))\;-\;t_{max}\;)}$$

with (nominal) parameters, $N_{max}$ and $t_{max}$. Note: this assumes the shower starts at $t = 0$ and it develops in *radiation lengths*.

# A New Approach to Shower Profiles (Ib)

- Gaisser-Hillas: We now replace $X_0$ using: $W \equiv X_{max} - X_0$ and define distances with respect to $X_{max}$: $\Delta = X - X_{max}$. Then the GH profile can be written as:

$$N(X; N_{max}, X_{max}, W, \lambda)_{GH} = N_{max}\,(1 + \frac{\Delta}{W})^{\frac{W}{\lambda}}\; e^{-\frac{\Delta}{\lambda}}$$

  This form emphasizes the role of a physical quantity, the distance from shower maximum, $\Delta$, in comparison to *e.g.* an unphysical quantity, $X - X_0$. Furthermore $\Delta$ is scaled by a parameter $W$ potentially resolving a tension between $X_{max}$ and $X_0$ in a parameter optimization to *fit* experimental shower profiles.

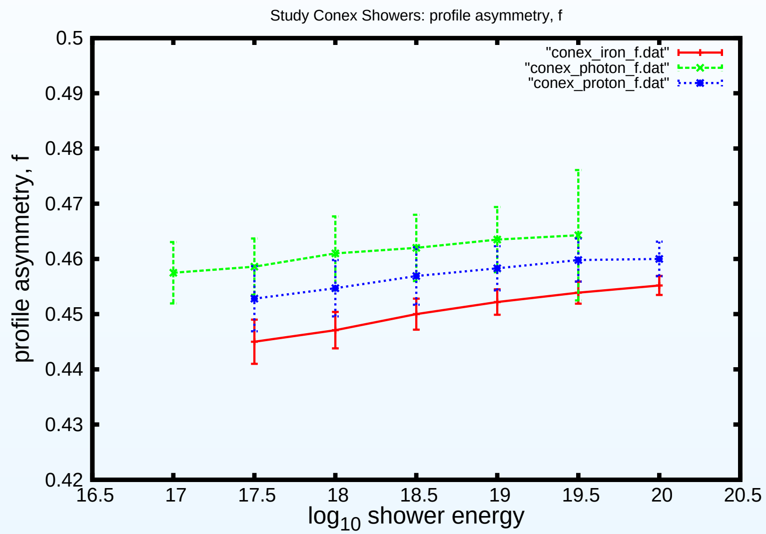
- Gaussian-in-Age: We introduce $X_0$ then replace it using: $\epsilon \equiv \frac{\Delta}{W}$ and $s(\epsilon) = \frac{1+\epsilon}{1+\epsilon/3}$:

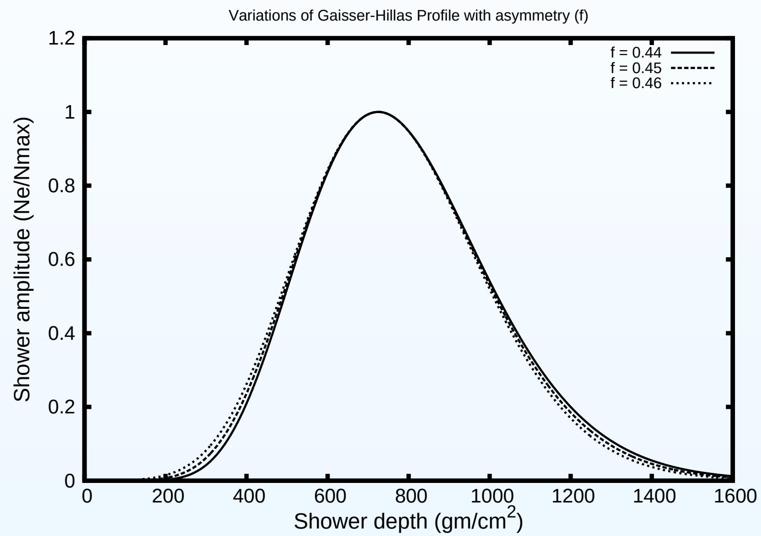
$$N(X; N_{max}, X_{max}, W, \sigma)_{GIA} = N_{max}\; e^{-2(\frac{\epsilon}{(3+\epsilon)\sigma})^2}$$

- Greisen: Similarly introducing $t_0$, converting to gm/cm$^2$ ($t = X/p_{36.7}$) ... :

$$N(X; N_{max}, X_{max}, W, p_{36.7})_{Greisen} = N_{max}\; e^{(\;\epsilon(1-\frac{3}{2}ln(s(\epsilon)))\;-\;\frac{3}{2}ln(s(\epsilon))\;)\;\frac{W}{p_{36.7}}}$$

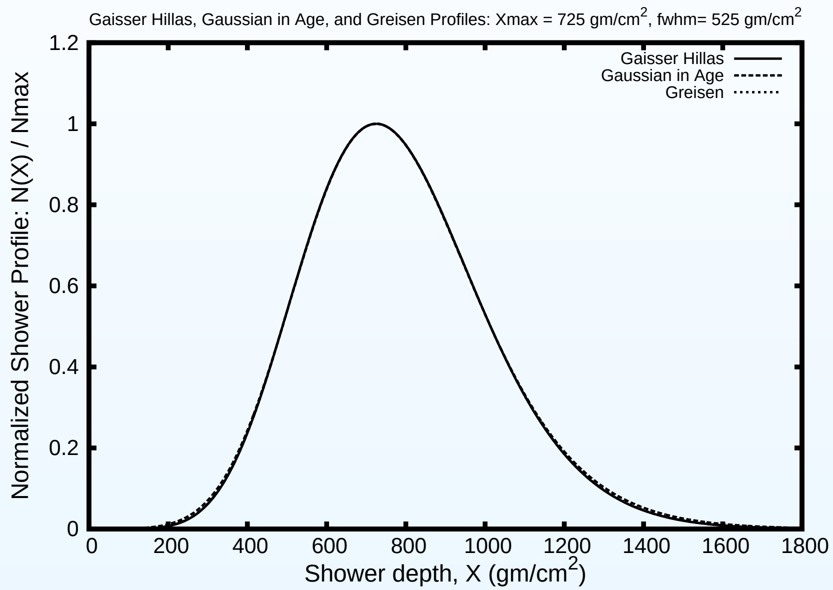
# A New Approach to Shower Profiles (II)

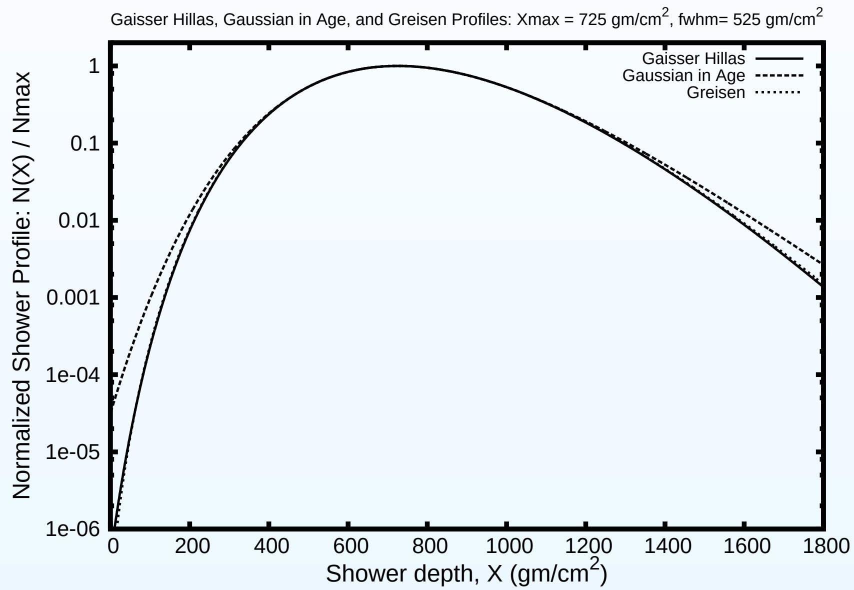



- **Left plot:** CONEX simulations suggest that the asymmetry parameter $f$ may provide some discrimination in primary composition
- **Note:** as the GIA "$\sigma$ parameter" and the GH "$(Xmax - X_0)/\lambda$ ratio" depend **only** on the asymmetry $f$, this echos the results of: V. Scherini et al (ICRC 2007) and S. Andringa et al (ICRC 2009)
- **Right plot:** But the effect is subtle. The GH shower profiles have $X_{max} = 725$ gm/cm$^2$, $fwhm = 525$ gm/cm$^2$ and three different values of asymmetry: $f = 0.44$, 0.45 and 0.46.

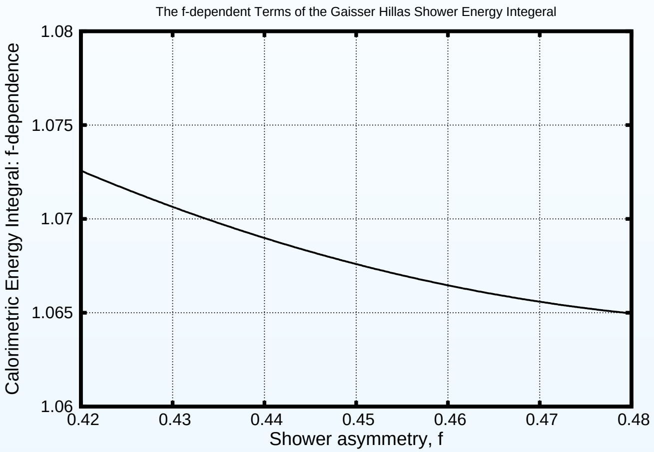
# What did we learn? (I)

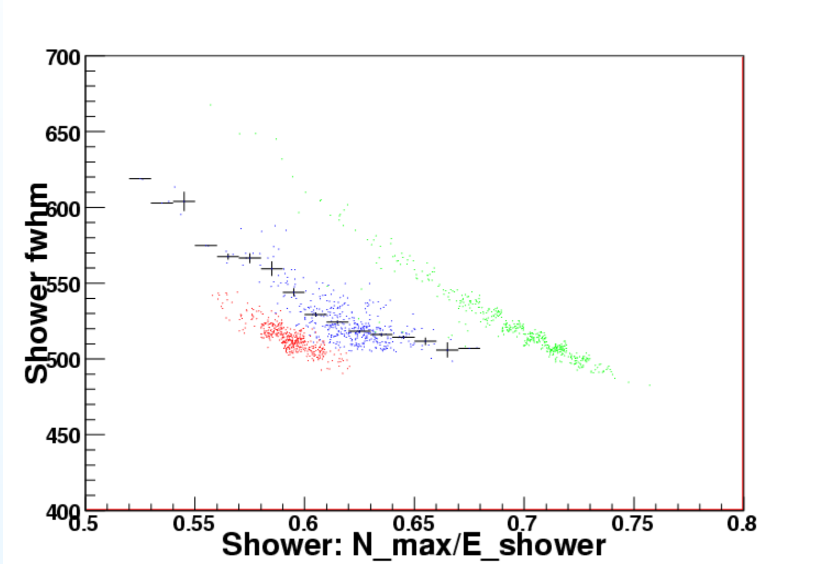



- **Left plot:** Shower profiles with the same $fwhm$ and asymmetry $f$ are almost indistinguishable
- **Right plot:** The GH and Greisen profiles are systematically below the GIA profile for shower depths well away from shower maximum. Thus shower calorimetric energies evaluated using the GIA function are $\sim 1\%$ larger than those evaluated using GH or Greisen forms.

# What did we learn? (II)



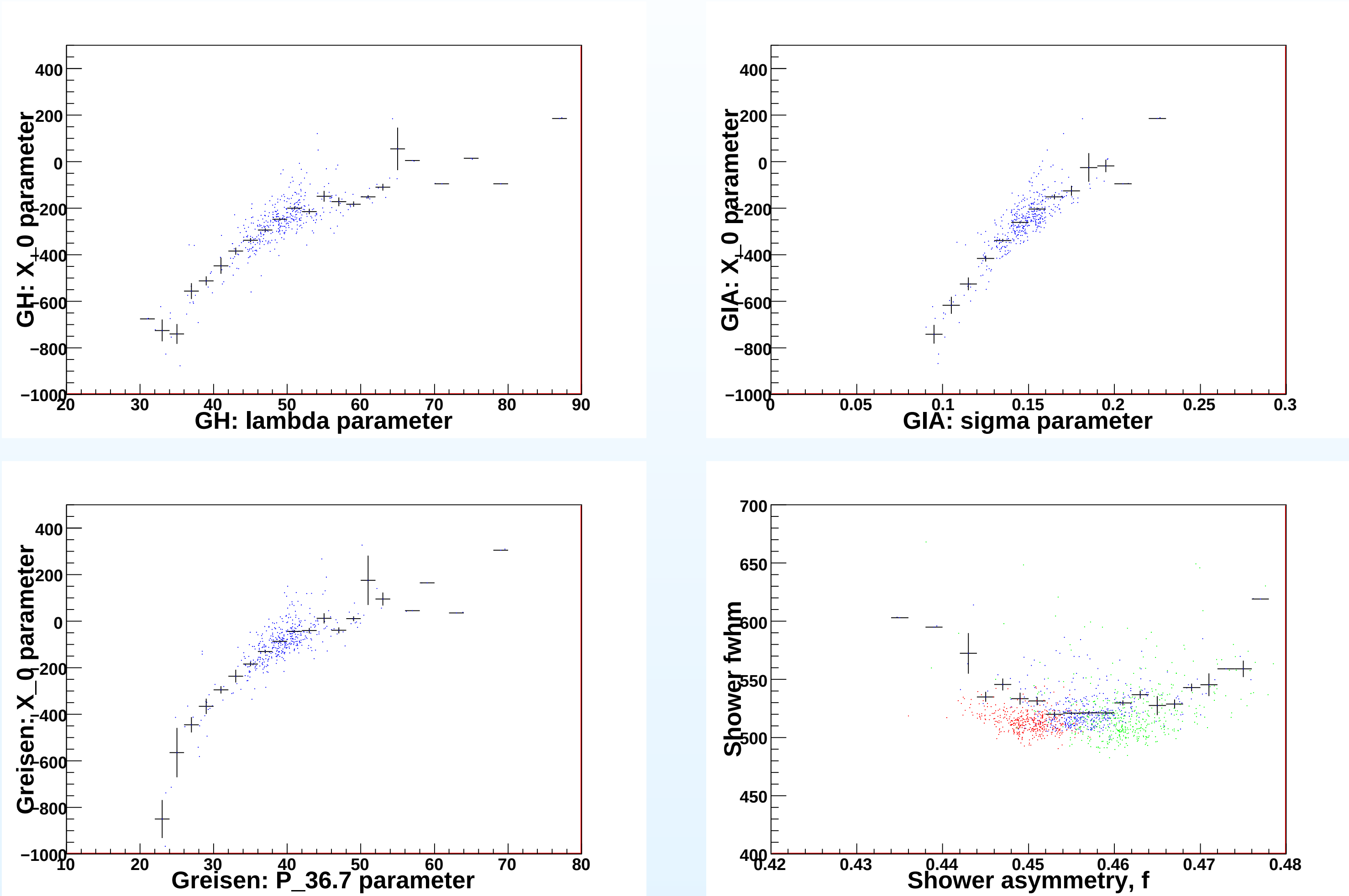




- The GH calorimetric shower energy is to a good approximation:

$$E^{calor}_{shower} \;=\; < dE/dx > \; N_{max}\; fwhm\; (\frac{\xi^{-(\xi+1)}\; e^{\xi}\; \Gamma(\xi+1)}{R(f)})$$

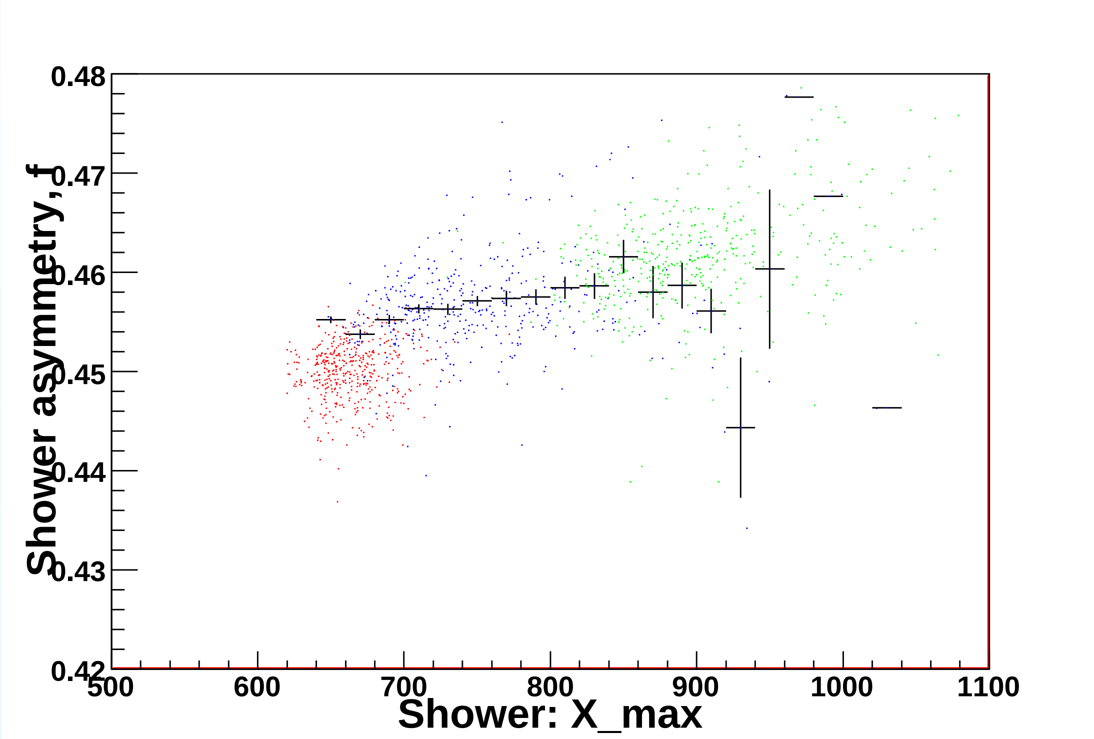
- **Left plot:** The asymmetry parameter $f$ dependence, terms in ( ), is small
- **Right plot:** Thus $E^{calor}_{shower} \propto N_{max}\; fwhm$; CONEX simulations are shown for proton, iron and photon showers at $10^{18.5}$eV

# What did we learn? (III)




- Shower ($fwhm$, $f$) parameters are less correlated than conventional parameters
- But correlated doesn't mean that 3 parameters are sufficient ...

# Composition and Exotics studies




- Plot of shower asymmetry $f$ VS $X_{max}$ for CONEX simulations of proton, iron, and photon showers near $10^{18.5}$eV
- As conventional showers have tails mostly to *larger* values of shower asymmetry $f$, typically associated with showers with *larger* values of $fwhm$, exotic studies are urged to search for showers with smaller values of shower asymmetry: *i.e.* more asymmetric showers!

# Conclusions: Shower Profile based on $fwhm$,$f$

- For profiles with the same ($fwhm$, $f$), the GH and Greisen shower profiles are essentially identical and systemically less than GIA for shower depths away from shower maximum.

- Of the three functions, GH is most convenient as the integral of the GH profile is an analytic function.

- Monte Carlo simulated air showers using CONEX, and parametrized in terms of the new parameters: ($fwhm, f$), have correlations (between those parameters) greatly reduced over the standard parametrizations *e.g.* Gaisser-Hillas parameters: $(X_0, \lambda)$.

- This allows shower profile reconstructions to add constraints (if needed) on the mostly uncorrelated parameters $fwhm$, $f$.

- While not a new result, the CONEX shower simulations suggest that the shower asymmetry parameter, $f$, may have some sensitivity to the incident cosmic ray particle type: *e.g.* p, C/N/O, Fe (but possibly not $\gamma$s).

- For all the details see: LANL arXiv:0909.4014 $\equiv$ J. Phys. G: Nucl. Part. Phys. 37 (2010) 025202.

## (II) **Anisotropy searches:** *A. Watson* *(Auger Celebration, 2005)*

The University of New Mexico

**Why are we still searching for the origin of cosmic rays ~ 95 years after the discovery?**

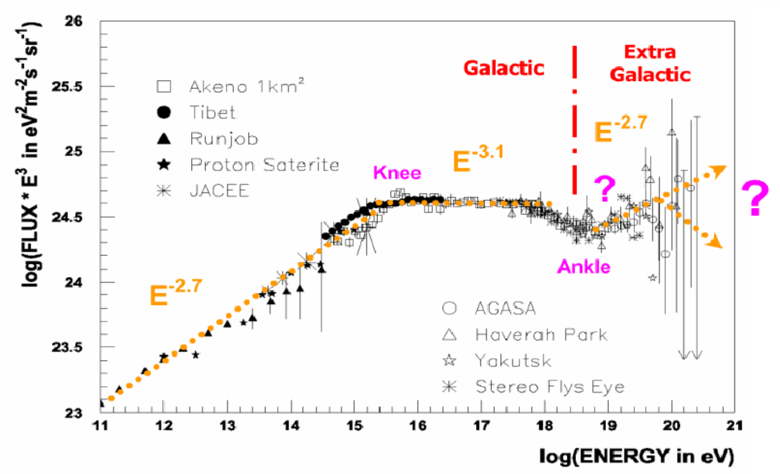
**Magnetic Fields are the problem:**

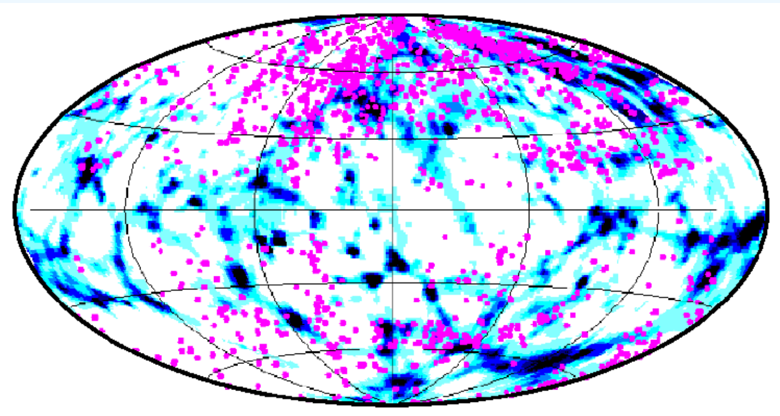
**While gamma-rays and neutrinos are 'blind' to magnetic fields, cosmic rays are charged particles, the nuclei of atoms.**

**Like the drunken man's walk!**

**BUT the highest energy particles are expected to be almost undeflected by the fields → cosmic ray astronomy.**

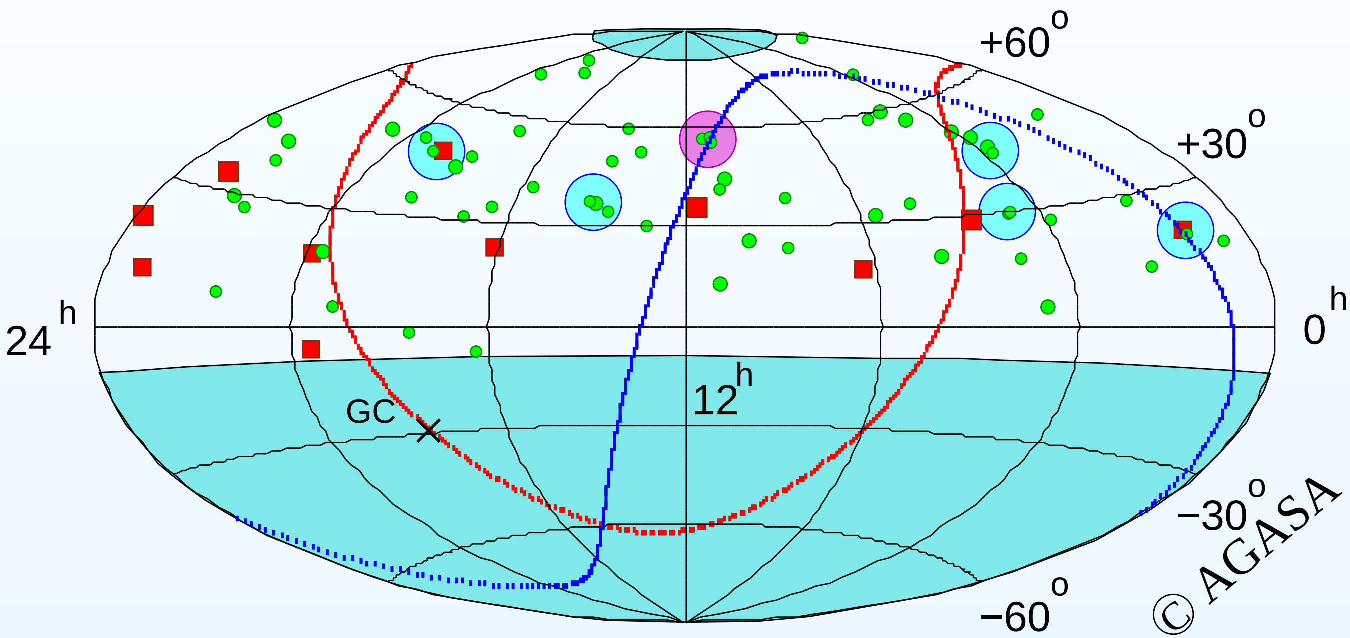
# Motivation for cosmic ray anisotropy:



- For several reasons, the highest energy CRs *e.g. with energies above the ankle*, Upper Figure, are probably from extra-galactic, astrophysical sources
- With a GZK cutoff, then the *highest energy CRs* should come from relatively nearby sources ...
- For nearby ($9 < R < 93$ Mpc), astrophysical sources, the universe is observed to be non-isotropic: Lower Figure
- Thus, baring magnetic field and/or composition surprises, we expect the arrival directions to show structure: *i.e.* **be anisotropic**
- And what is the best way to search for anisotropy signal(s): *clusters of CRs, CR correlations with astrophysical catalogs, non-isotropy in CR arrival directions, ...* consistent with small (low statistics) data samples?

# Experimental examples: AGASA




- If sources are *bright* we expect to see multiple cosmic rays/source
- AGASA reported 5 doublets and 1 triplet few-degree sized event-*clusters*
- With larger exposure, HiRes stereo data have **not** verified the AGASA result
- However if sources are *faint*: then searching for (cross-) correlations between *candidate* sources and CRs may show a signal

# Experimental examples: AGASA, Yakutsk, HiRes

Testing the correlations between ultra-high-energy cosmic rays and BL Lac type objects with HiRes stereoscopic data

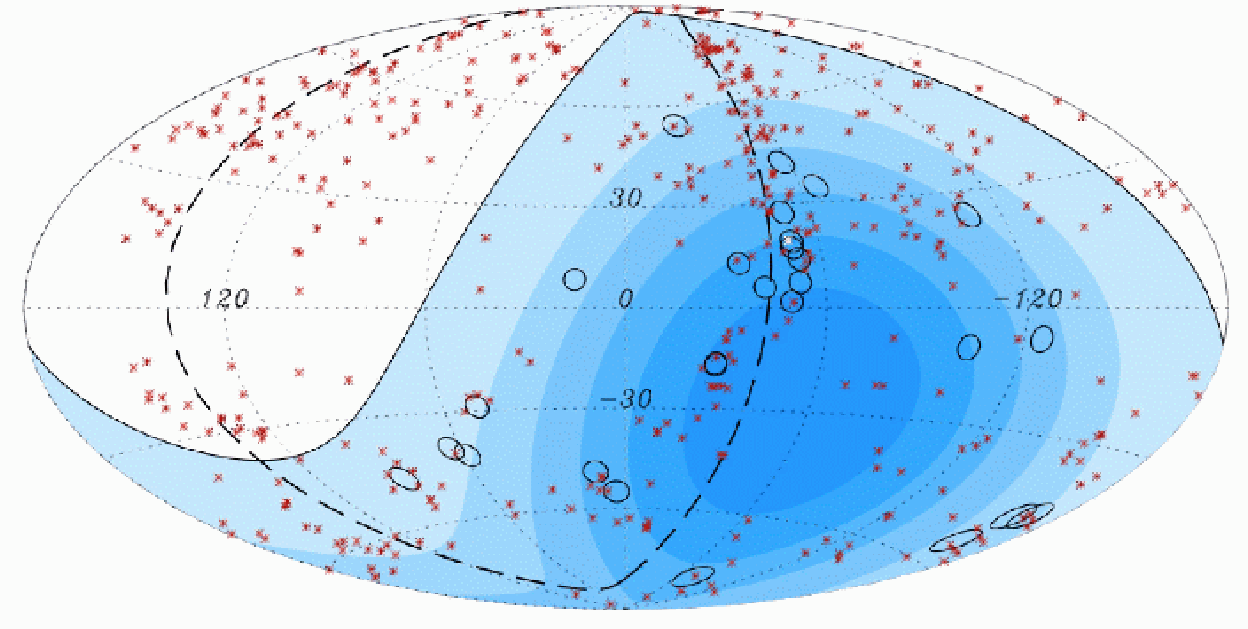
D. S. Gorbunov[a 1)], P. G. Tinyakov[b,a], I. I. Tkachev[c,a], S. V. Troitsky[a]

[a] Institute for Nuclear Research of the Russian Academy of Sciences, 60th October Anniversary Prospect 7a, 117312, Moscow, Russia;
[b] Service de Physique Théorique, CP 225, Université Libre de Bruxelles, B–1050, Brussels, Belgium;
[c] CERN Theory Division, CH-1211 Geneva 23, Switzerland.

Previously suggested correlations of BL Lac type objects with the arrival directions of the ultra-high-energy cosmic ray primaries are tested by making use of the HiRes stereoscopic data. The results of the study support the conclusion that BL Lacs may be the cosmic ray sources and suggest the presence of a small (a few percent) fraction of neutral primaries at $E > 10^{19}$ eV.

- Popular *candidate* astrophysical sources for UHECRs include active galactic nuclei (AGNs) and gamma ray bursts (GRBs) ... but we do not know!
- While some correlations have been found, confirming their significance with low statistics data was difficult.
- Higher statistics Auger data are inconsistent with the BL Lac:CR correlation!

# Experimental examples: Auger

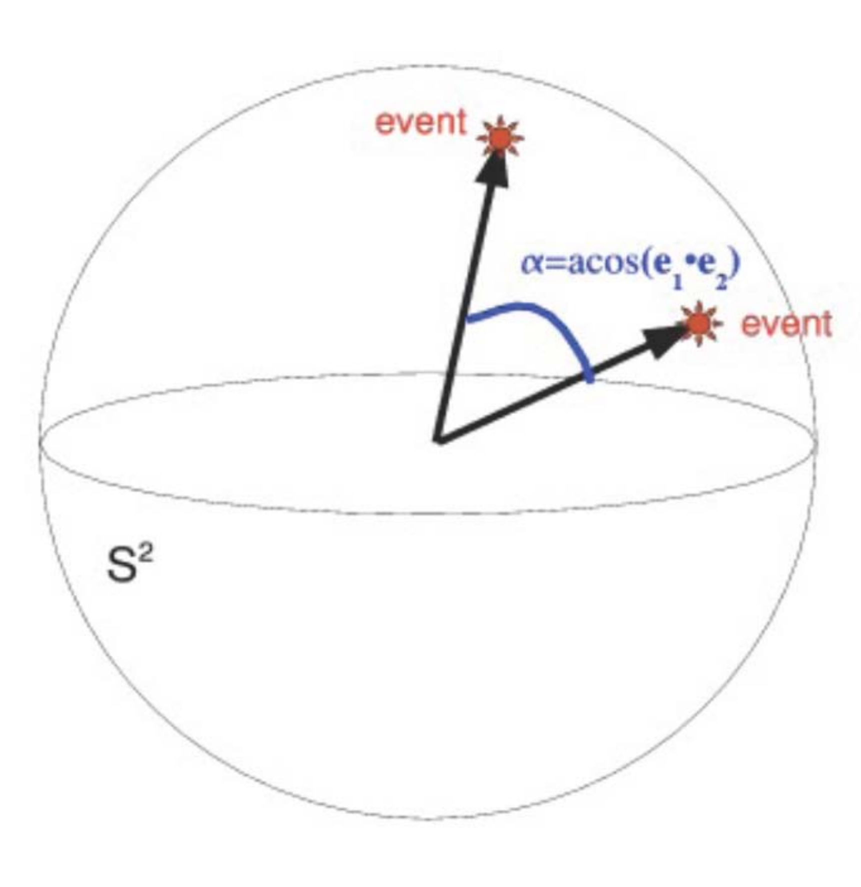



- The most compelling observational evidence consistent with astrophysical expectations of anisotropy is arguably the 27 events with energy $> 57$ EeV observed by Auger.
- At a minimum, the Véron catalog: AGN maximum redshift and correlation angle, defines a limited area (effectively 21%) of the sky. Thus the Véron catalog AGN:CR correlation signal is evidence for a non-isotropic flux of CRs that is enhanced near known extra-galactic objects.

# Catalog independent methods:

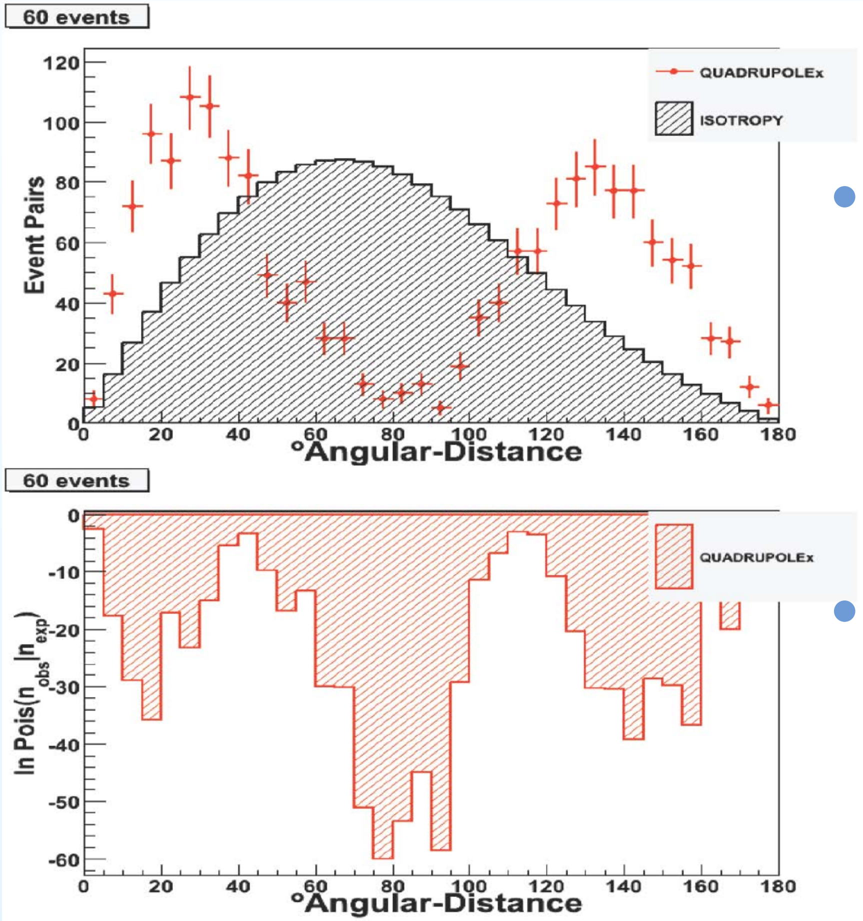
- Catalog dependent (cross-correlation) studies are not without issues: *e.g.* penalty factors for scans over different catalogs, issues related to *brightness limited* catalogs, and/or the need to restrict the data to match the limited sky coverage of individual catalogs.
- However with limited statistics, catalog independent (auto-correlation) methods are intrinsically less sensitive than (any given) catalog dependent study.
- Thus there is a need to identify and/or develop more effective (catalog independent) methods.
- We have studied two catalog independent analysis, C.I.A., techniques:
  1. a binned two-point (2-point) angular correlation method for all pairs of CR events
  2. a new, (binned), three point (3-point) method that uses a shape and strength parameter for all triples of CR events.
- Our paper is available at: arXiv:0905.4488 $\equiv$ J.Phys.**G36** 115203 (2009)

## A 2-point method:




- Two points on the sphere define an angle.
- Use the set of angles between all pairs of CRs.
- Compare the observed distribution VS isotropic expectation (for the same size, Monte Carlo, data sample).
- Use a Pseudo-Likelihood test statistic
- Thus our 2-point analysis is for pedagogy: a “known” to be easily compared with our 3-point method, an “unknown”!

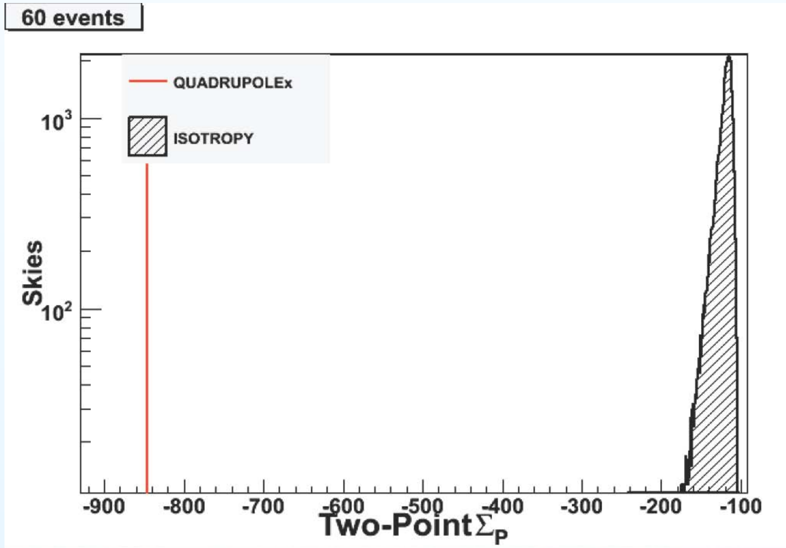
# A 2-point method ( *toy example - I* ):




- *Toy* CR Monte Carlo (60 event) data set was generated with a quadrupole distribution on the sky.
- Upper plot: The distribution of angles between all pairs of CRs are the red points with error bars; the gray histogram is the isotropy expectation.
- Lower plot: The probability for observing $n_{obs}$ doublets in the $i^{th}$ bin given that we expect $n_{exp}$, is approximated by a Poisson distribution:

$$P_i(n_{obs}|n_{exp}) = n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}/n_{obs}!$$

## A 2-point method ( *toy example - II* ):




- We then compute a pseudo-log-likelihood:

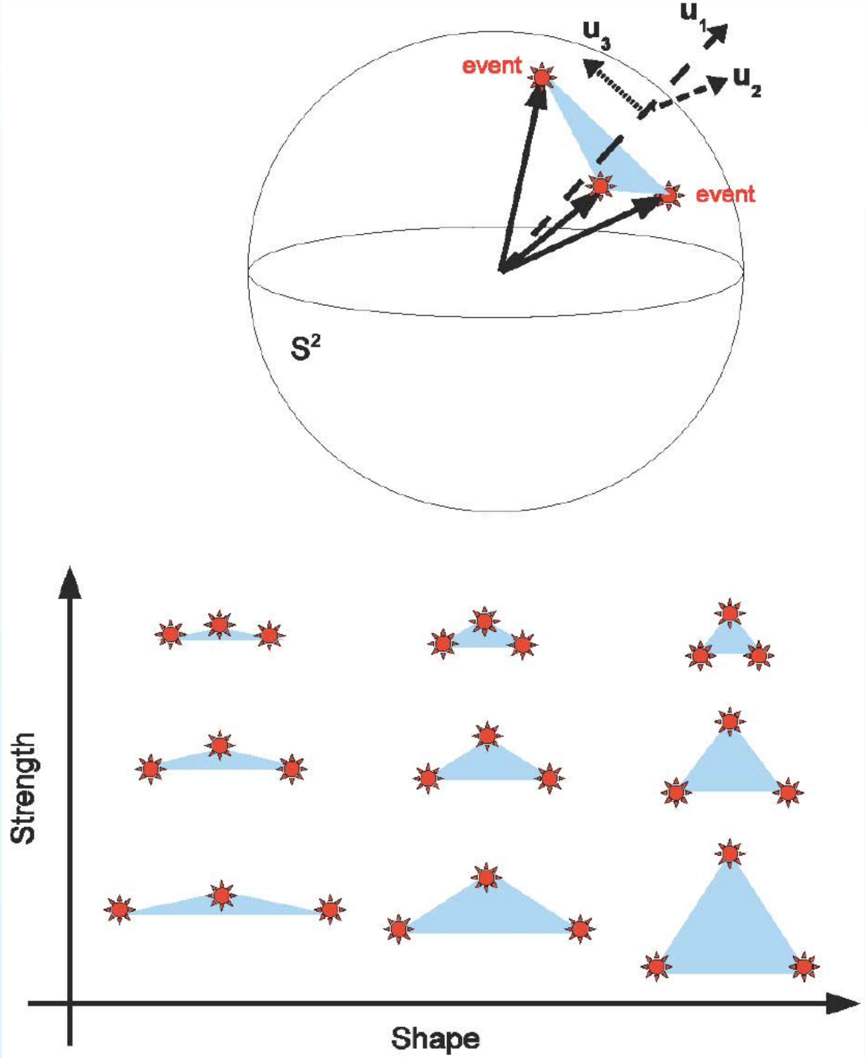
  $\Sigma_P = \Sigma_{i=1}^{N_{bins}} \ln \mathrm{P}_i(n_{obs}|n_{exp})$

  for the *toy* CR data set.

- The distribution of pseudo-log-likelihoods for a large number of *equivalent* isotropic data sets (typically 20,000) is also plotted (hatched histogram).

- Quantitatively: the significance, $P$, is the fraction of Monte Carlo *equivalent* isotropic data sets to the left of the red line.

# A 3-point method:




- Three points on the sphere
- Use the eigenvalues of rotation matrix: $\tau_1 \geq \tau_2 \geq \tau_3$, and $\tau_1 + \tau_2 + \tau_3 = 1$, thus there are only two free parameters
- Define the:

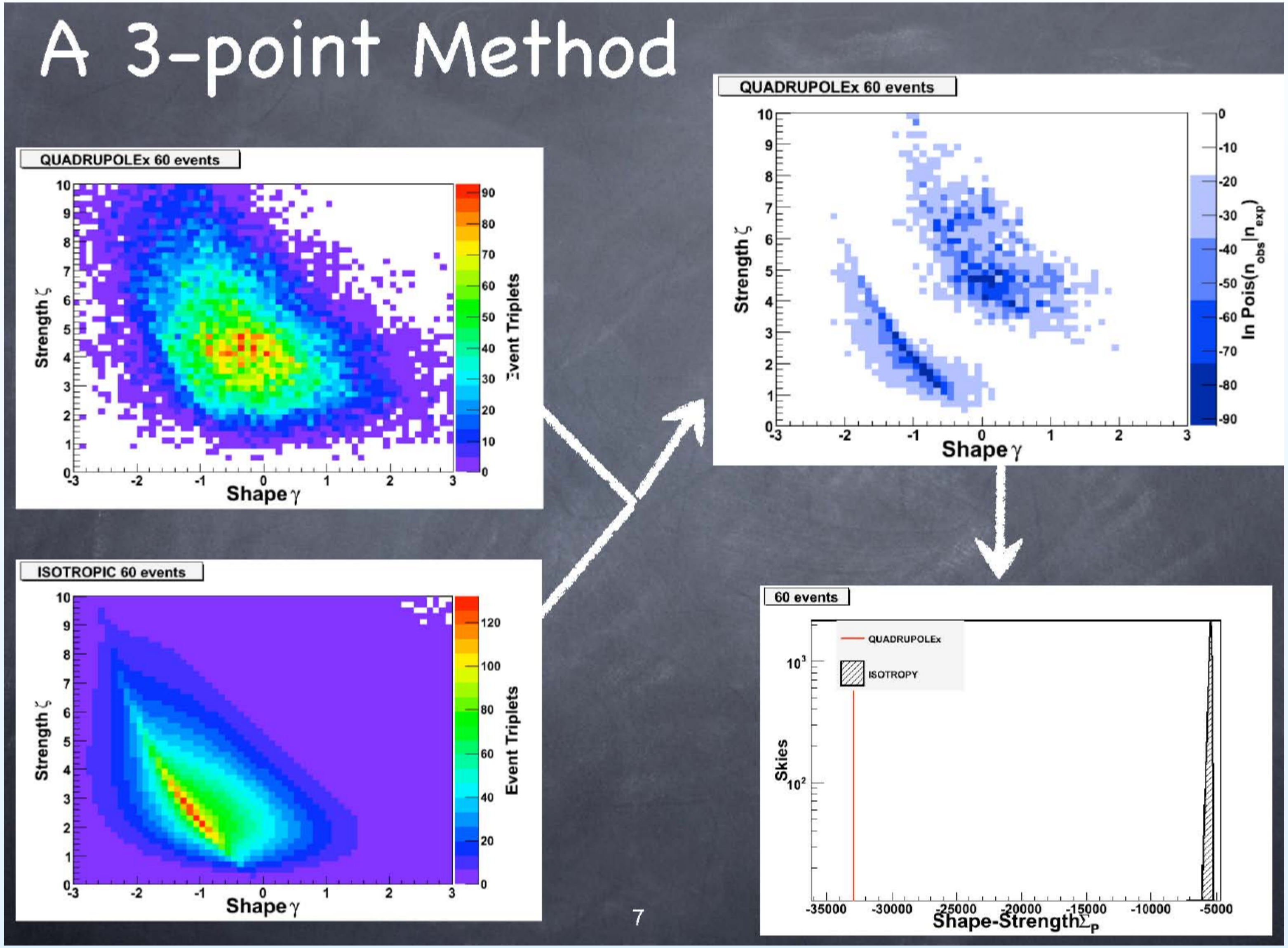
  Shape: $\gamma = log( \frac{log(\tau_1/\tau_2)}{log(\tau_2/\tau_3)} )$

  as the shape increases from $-\infty$ to $+\infty$ the triples are *less elongated*.

  Strength: $\zeta = log( \tau_1/\tau_3 )$

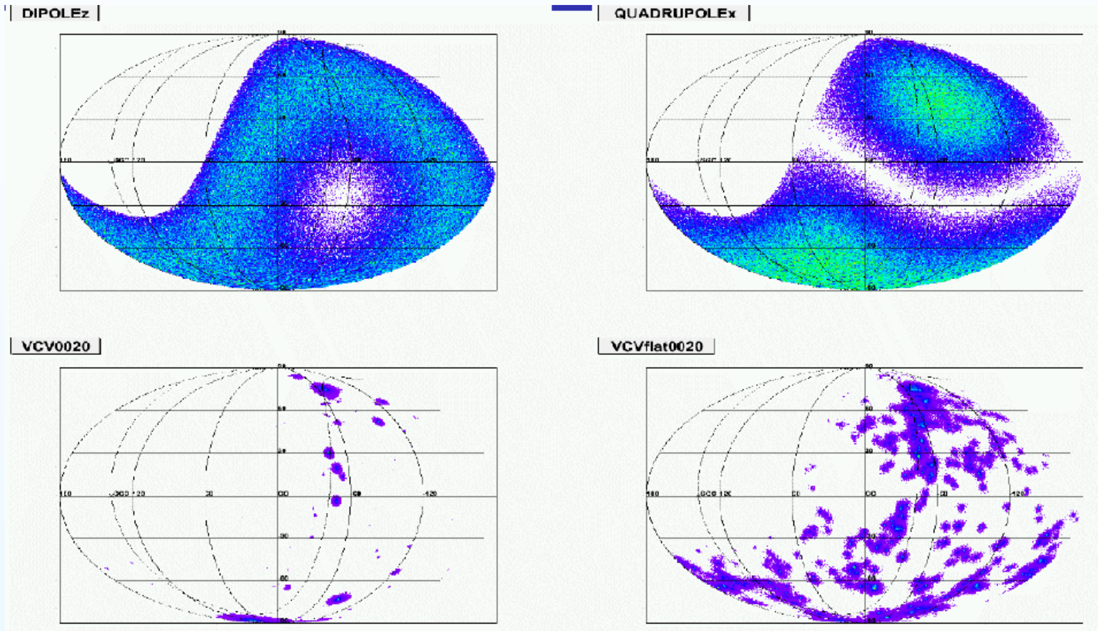
  as the strength increases the CR events are more *concentrated*.

# A 3-point method ( *toy example* ):



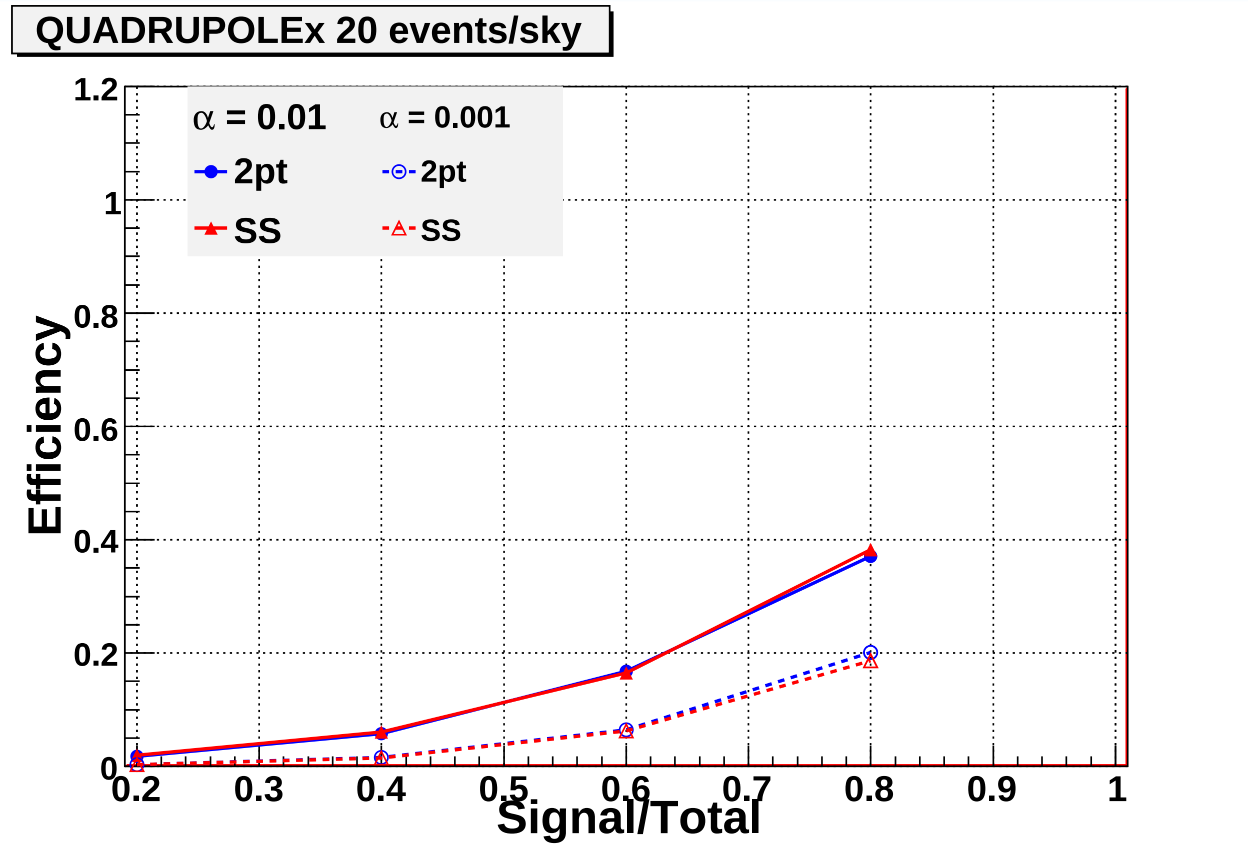
# Catalog independent analyses (I)

The University of New Mexico




- To study the sensitivity of our *metrics*, we used several *mock* anisotropic models.
- All studies required: $P <$ Type I error $\alpha < 1\%$ (or 0.1%), so anisotropy is distinguishable from isotropy. For good detection efficiency the: Type II error (determined via simulation) should be: $\beta < 10\%$, *i.e.* the efficiency (*power*) is $1 - \beta$.
- Studies were done varying: data set size and the fraction of anisotropic events.
- Four of the *mock* distribution are shown (in galactic coordinates) weighted by the acceptance of the Auger Southern Observatory.

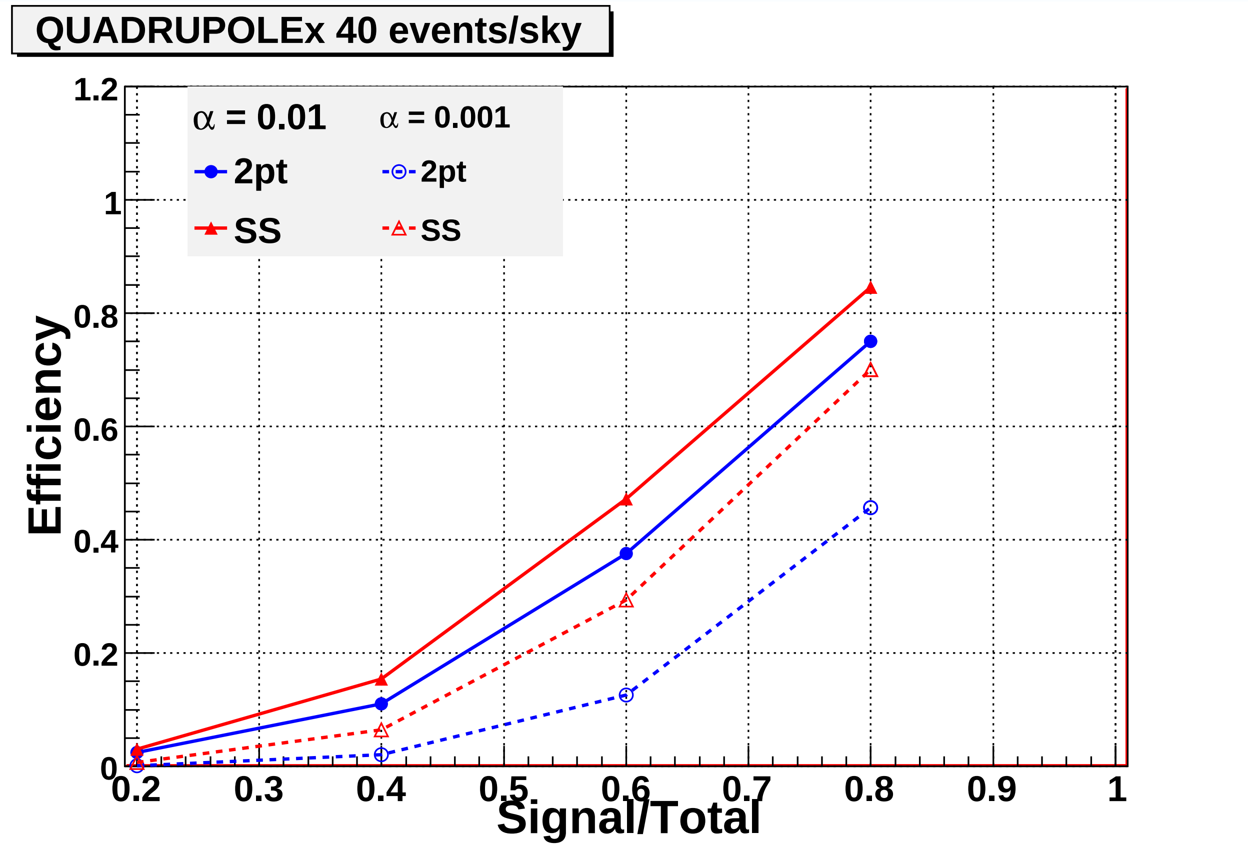
# Catalog independent analyses (II)




How do source detection *efficiencies* vary with source purity (*i.e.* signal/total) and number of CR events?

20 events

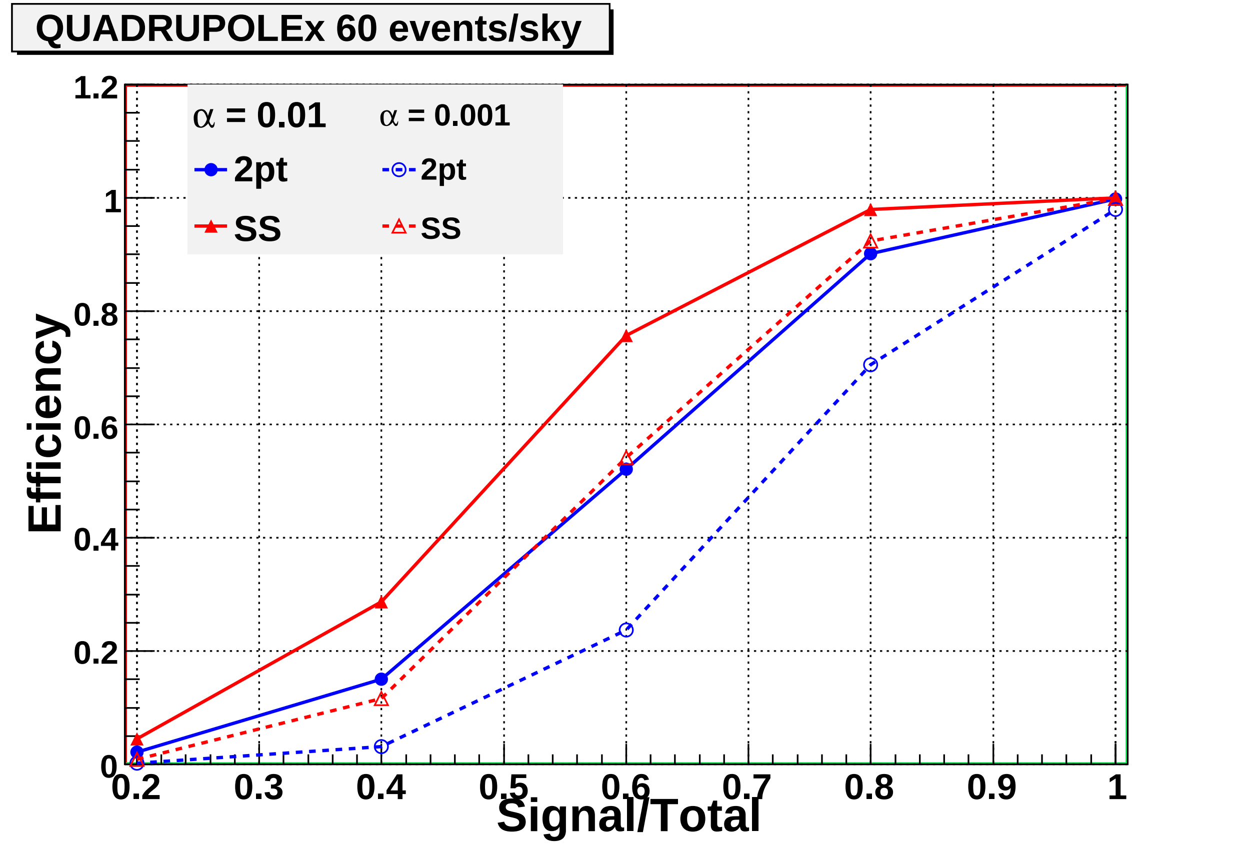
# Catalog independent analyses (III)




How do source detection *efficiencies* vary with source purity (*i.e.* signal/total) and number of CR events?

40 events

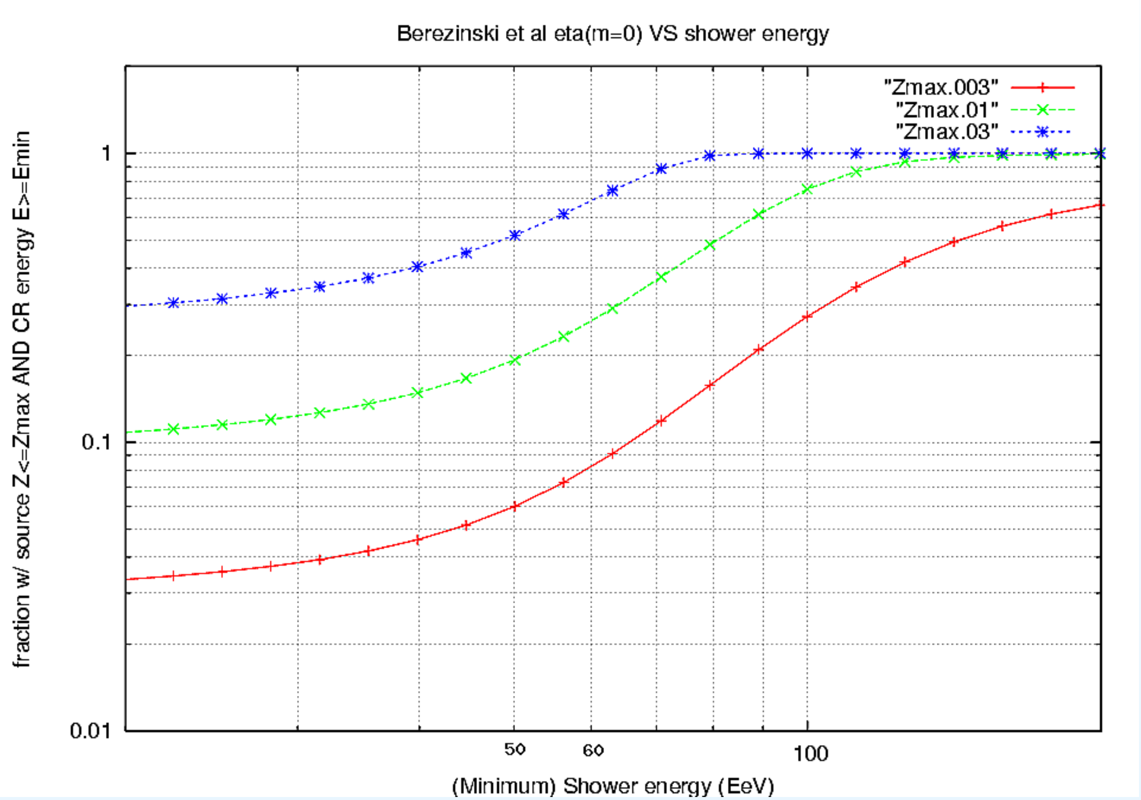
# Catalog independent analyses (IV)




How do source detection *efficiencies* vary with source purity (*i.e.* signal/total) and number of CR events?

60 events

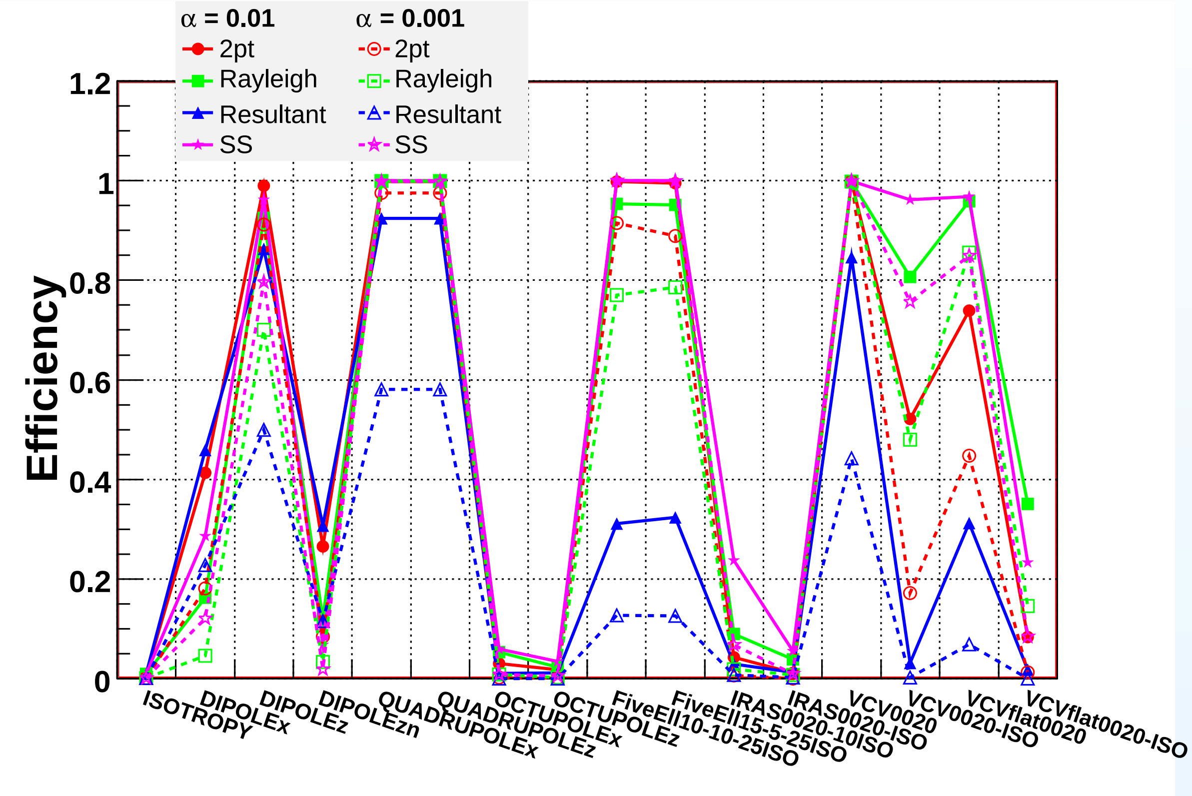
# Catalog independent analyses (V-a)




There is a tension between *statistics* and *fraction of events from nearby: $z \leq 0.02$ sources*:

- catalog independent techniques profit from more events
- yet assuming CR protons, GZK models suggest that for a threshold energy as low as 60 EeV the fraction of CRs from nearby sources is $\sim 50\%$ ... which negatively impacts detectability!

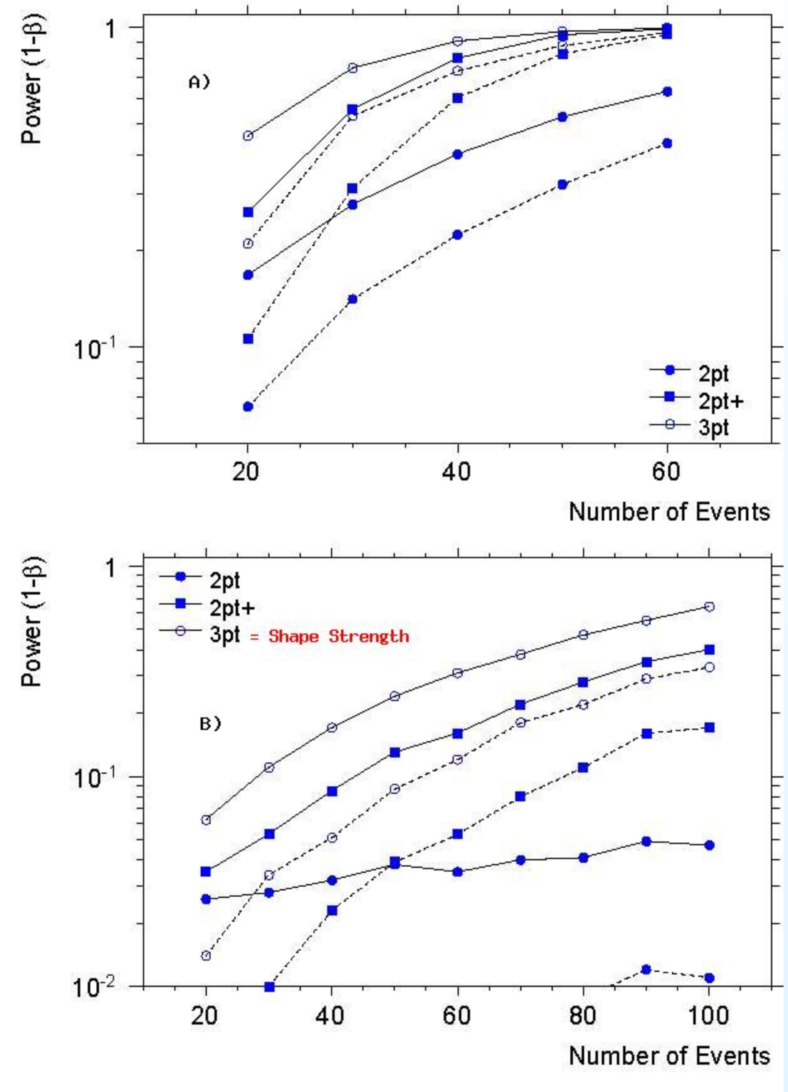
# Catalog independent analyses (V-b)




Based on simulated samples (*ie mock data*) from hypothetical sources, we find:

- some source (distributions) can be identified (at the 1% or 0.1% confidence level) with 60 events and some cannot!
- the sense is that many more than 60 events may be needed for a robust identification of an anisotropic signal in the highest energy cosmic rays.

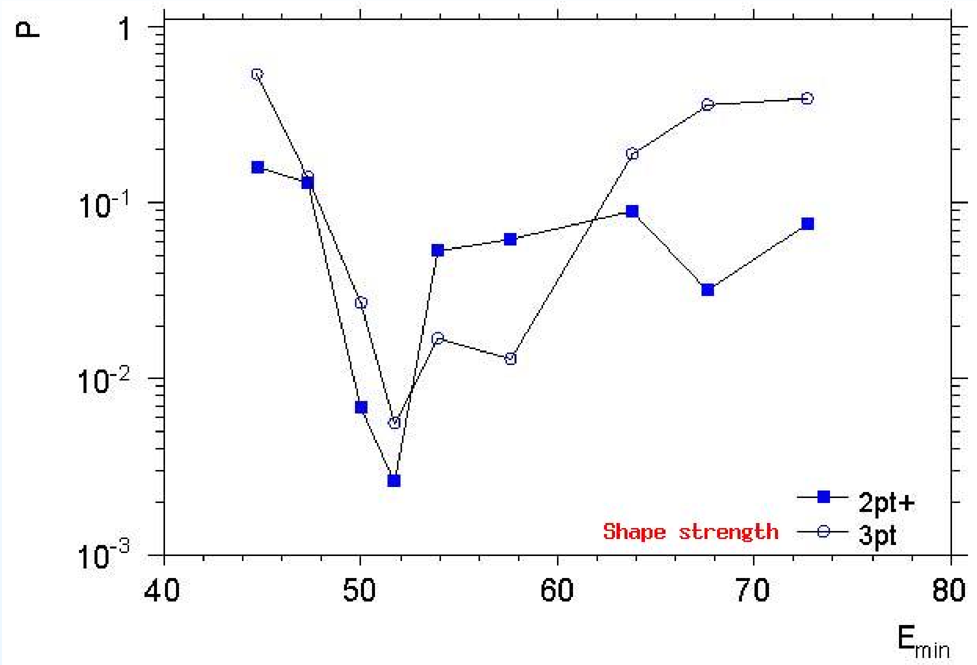
# C.I.A. application to Auger data (I)




- Top plot: Monte Carlo study of *mock* data from astrophysically motivated sources similar to Vernon catalog AGNs. The plot shows the power of three catalog independent analyses: 2pt, 2pt+ and 3pt as a function of the number of CR events. The 2pt and 3pt are the methods presented earlier.
- Solid lines are for $\alpha = 1\%$, dashed lines are for $\alpha = 0.1\%$.
- Bottom plot: The equivalent plots assuming 50% isotropic background.

# C.I.A. application to Auger data (II)

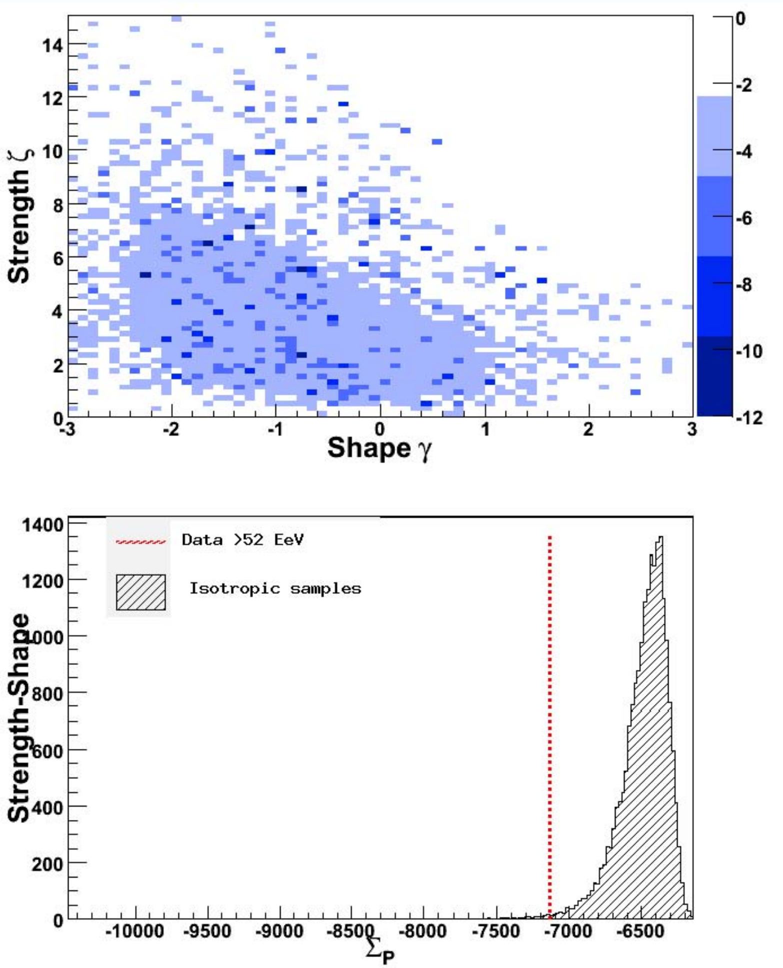
The University of New Mexico




Scan of Auger data: January 1, 2004 to March 31, 2009 starting with the 20 highest energy events then in steps of 10 events to 100 events.

- The vertical axis is the probability, $P$, for the data to be a realization of an isotropic source distribution. The minimum values are: $P = 0.26\%$ (2pt+) and $P = 0.56\%$ (3pt) for $E_{min} \approx 52$EeV (rather similar to our AGN:CR correlation result).
- NB: bins are correlated and no *scan penalty* correction has been made in reporting, $P$.

# C.I.A. application to Auger data (III)



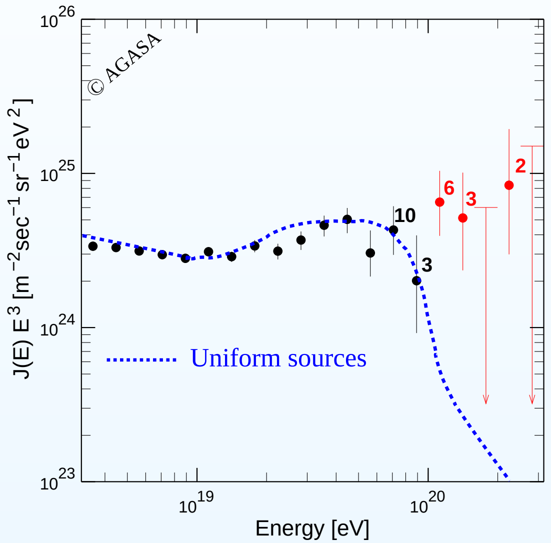

- Top plot: Plot of the natural-log of the Poisson probability to observe $n_{obs}$ shape-strength triples given $n_{exp}$ assuming an isotropic distribution (in shades of blue).

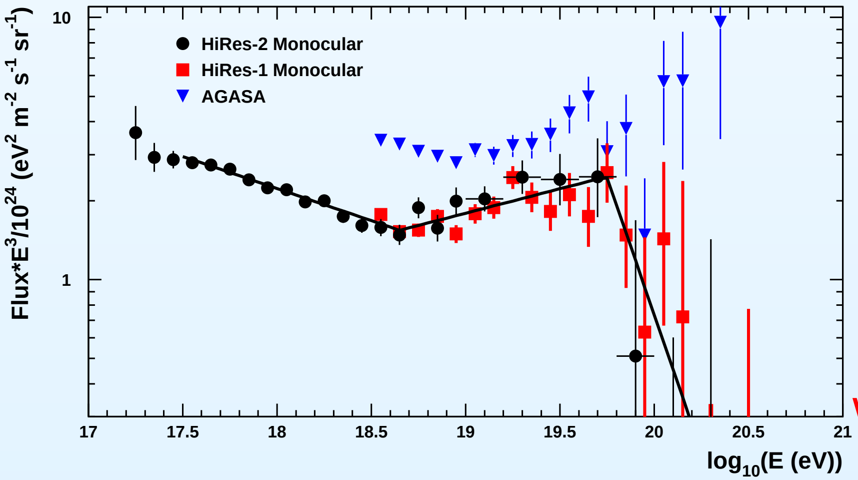
- Bottom plot: The distribution of Pseudo-log-likelihoods for 20,000 (equivalent) isotropic data sets is shown *hatched*. The red line is the Pseudo-log-likelihood for the 70 events above 52 EeV. The significance, $P$, is the fraction of Monte Carlo (equivalent) isotropic data sets to the left of the data.

# Anisotropy searches: Conclusions

- We have illustrated a new, 3-point (shape-strength), C.I.A. method for detecting anisotropy in spherical data sets that is powerful for small numbers of events.
- Studies were done with many *mock* signals: signal type, number of events and signal dilution all dramatically effect detectability.
- Number of events and signal dilution:
  - If “lucky” then $\sim 60$ events with $\gtrsim 70\%$ signal are detectable
  - If not then many more events are needed
  - Tantalizing first C.I.A. results from Auger Southern Observatory
- Experimentally:
  - Many more events are likely needed for robust C.I.A. identification of anisotropy, (*i.e.* not “lucky”) detection.
  - Many more events + GZK-cutoff = very large detector

# (III) What evidence for a GZK-cutoff?




**If the highest energy cosmic rays are non-isotropic, this is strong circumstantial evidence for a GZK cutoff!**

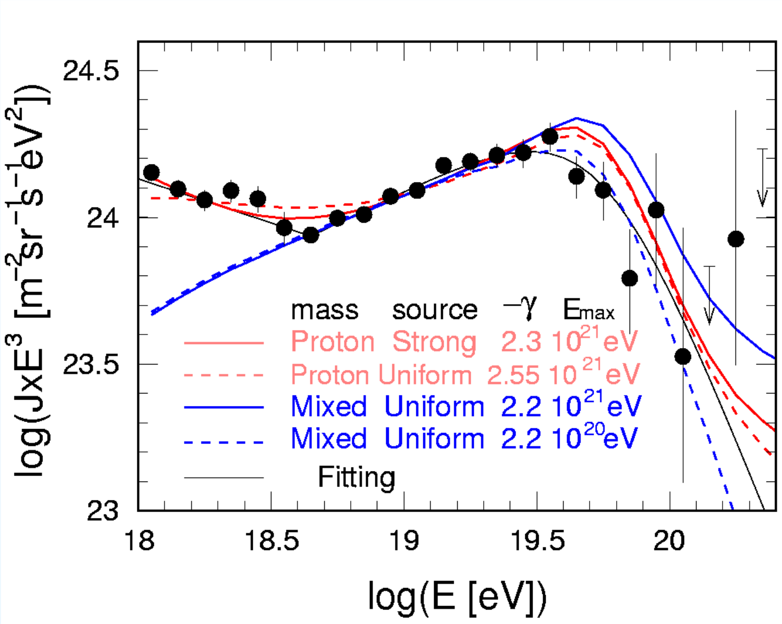
(top) AGASA spectrum

(bottom) HiRes spectrum Phys. Lett. **B619** 271 (2005) and [astro-ph/0703099]:

- The ankle shows up clearly at $4.5 \times 10^{18}$ eV ($\log_{10} E = 18.65$).
- The spectrum steepens again at $5.6 \times 10^{19}$ eV ($\log_{10} E = 19.75$).
- The fall-off of the HiRes spectrum above $10^{19.8}$ eV is evidence for the GZK cutoff.

What does Auger observe? And does Auger see a *cutoff in the UHECR spectrum?*

# Our Approach: **Measuring Flux Suppression**




**(Left plot) The Auger Flux** $\times E^3$ (ICRC'07). The suppression is "obvious" but quantification should be done carefully.

Our *eyes* like the binned-$E^3$ flux plot but their *statistical estimators* have some draw backs.

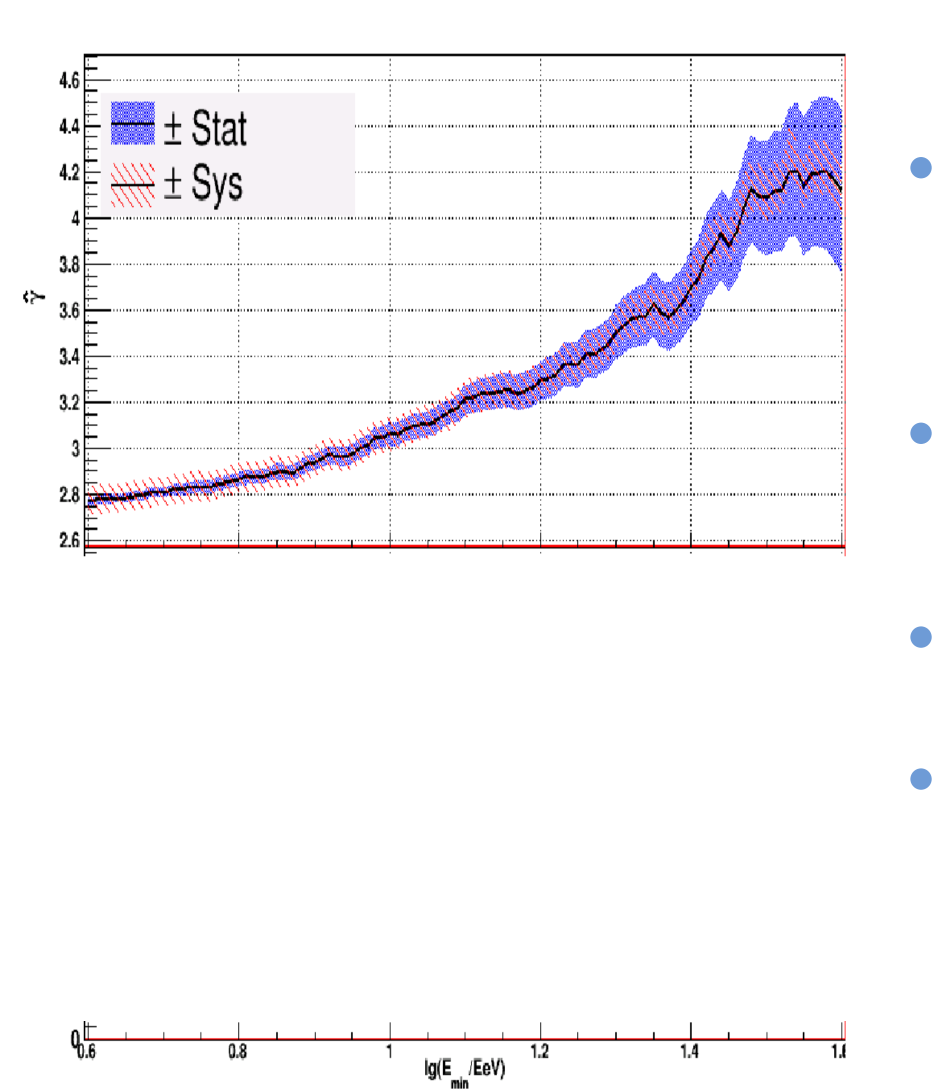
J. (Doug) Hague has provided two *statistical estimators* that are of general interest!

*See: arXiv:0710.3600 and astro-ph/0610865* $\equiv$ *Astropart.Phys.27:455-464 (2007)*

## We choose the following:

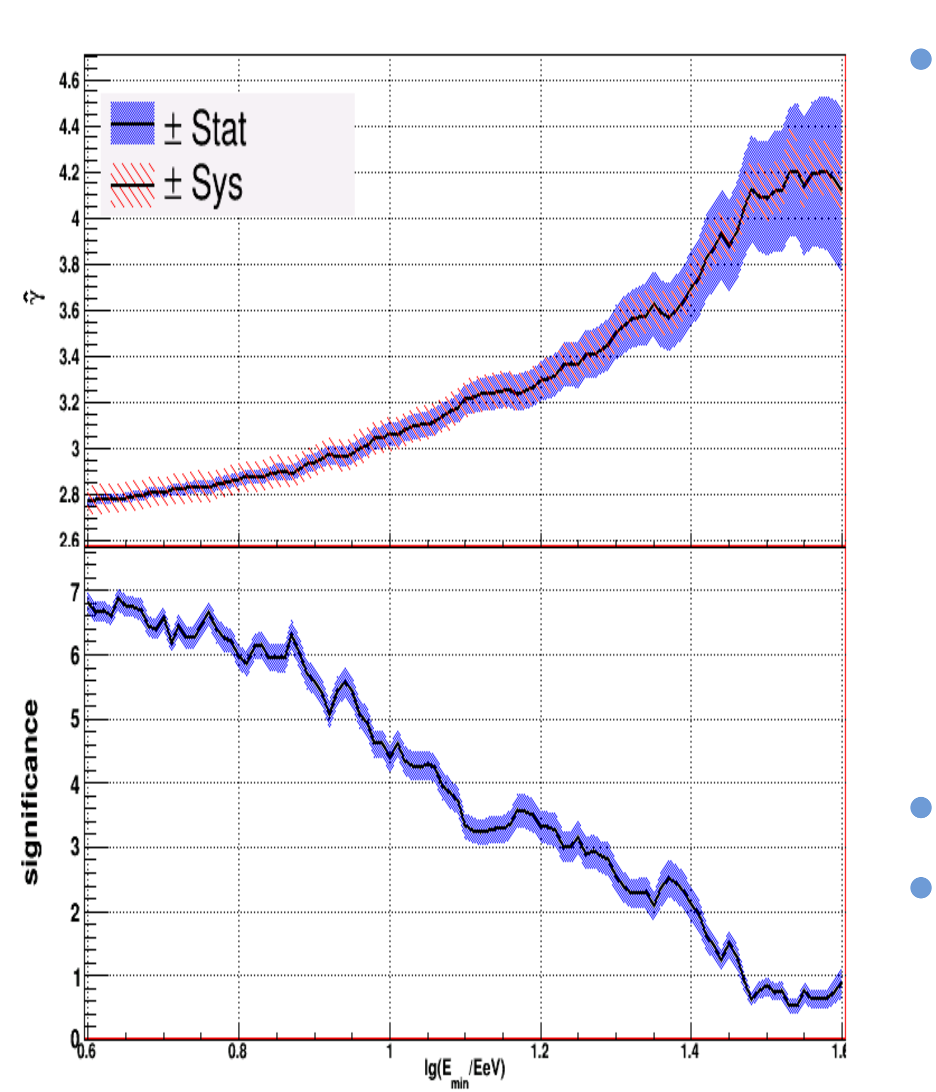
- Un-binned estimators as they are less correlated, more precise and more accurate.
- The Tail-Power (TP) statistic (which is identically zero for a pure power-law) can reject **non-**pure power-laws. It is (nearly) independent of the measured spectral index $\gamma$ and can discriminate tail suppression from tail enhancement.
- If a characteristic cutoff energy is desired, then a Likelihood Ratio Test has only a weak dependence on $\gamma$.

## **Hague** *Flux Suppression:* **The Hill plot**




- For each $E_{min}$, we determine the un-binned estimate of the pure power-law *spectral index* $\gamma$ (by maximizing the likelihood: **Top plot**).
- The systematic (energy) errors dominate for low $E_{min}$ but statistical errors dominate at large $E_{min}$.
- The index **increases** as the energy **increases.**
- There is suppression (*i.e.* the slope increases)! But how do we determine the significance?

# Hague *Flux Suppression:* The TP-statistic




- The TP-statistic ($\tau$) can discriminate between flux suppression (increasing slope with energy) and enhancement (decreasing slope with energy):

$$\hat{\tau}(E_{min}) = \hat{\nu}_1^2(E_{min}) - \frac{1}{2}\hat{\nu}_2(E_{min})$$

where:

$$\hat{\nu}_n(E_{min}) = \frac{1}{N_>} \sum_{E_i > E_{min}} \ln^n \frac{E_i}{E_{min}}$$

- It is (nearly) *independent* of $\gamma$.
- We can directly measure the *significance* in standard deviations of the flux suppression **(Bottom plot)**

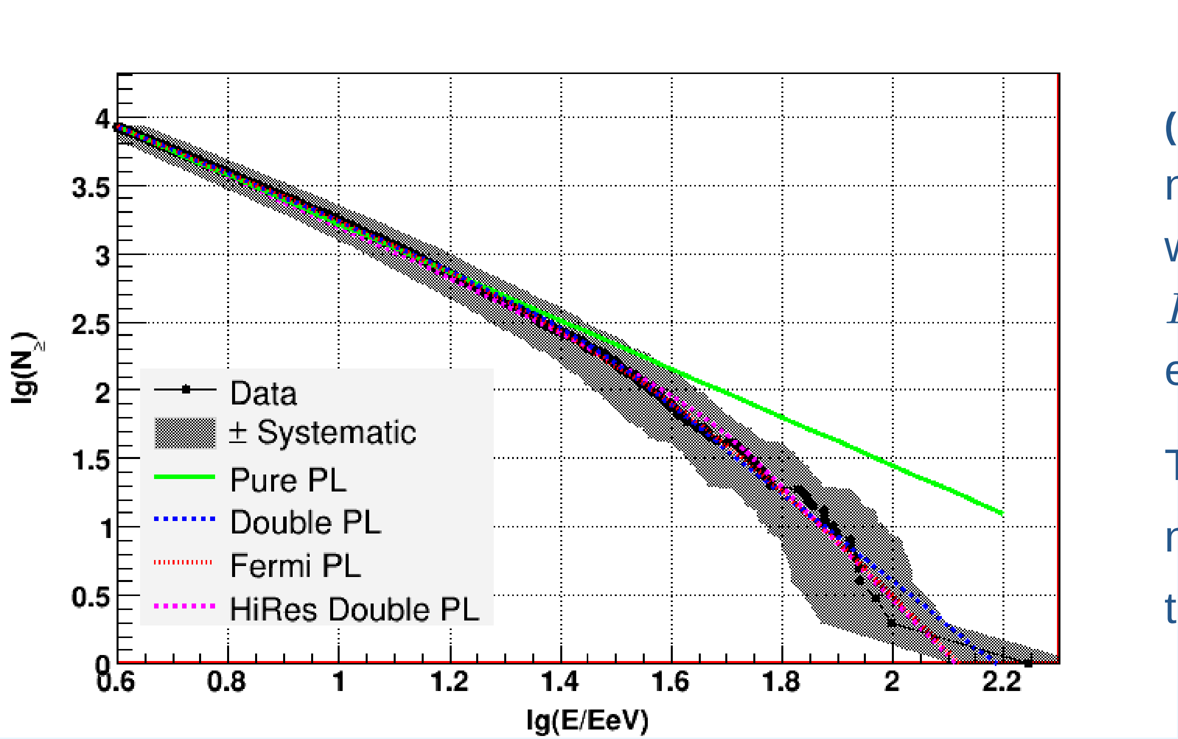
# Hague *Flux Suppression:* Fitted Models (I)

- We study three models "$f$" with parameters "$\theta = \{\theta_0, \theta_1, \ldots\}$":
  - The *pure* power-law: $\theta = \{E_{min}, \gamma\}$
  - and two models with tail *suppression*:
    1. the double power-law: $\theta = \{E_{min}, \gamma, E_b, \delta\}$
    2. a Fermi-like power-law: $\theta = \{E_{min}, \gamma, E_{1/2}, w_c\}$
- Parameters ( $\theta$ ) maximize the log-likelihood:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \ln f(E_i|\theta)$$

- Systematic (CR event) energy uncertainties are incorporated by shifting all event energies and then re-maximize the likelihood.
- Statistical (CR event) energy errors and acceptance information can be taken into account by the appropriate convolution.

# Hague *Flux Suppression:* Fitted Models (II)




**(Left plot)** A log-log plot of the number of (Auger) events with energy greater than $E_{min}$ *vs* event minimum energy ($E_{min}$).

The vertical axis is "one minus the (cumulative distribution function) CDF."

We plot:

- each event energy (with its systematic errors shown in gray)
- the three models; pure power-law, double power-law and Fermi-like power-law
- the reported HiRes double power-law (normalized to the Auger flux).

**Next:** we must now quantify the flux suppression.

# **Hague** *Flux Suppression:* **Likelihood Ratio**

- We can use the likelihoods to discriminate models. The Likelihood Ratio is:

$$\mathcal{R} = \frac{\mathcal{L}(\mathsf{data} \mid \mathsf{suppressed\ model\ hyp.})}{\mathcal{L}(\mathsf{data} \mid \mathsf{pure\ power\ law\ hyp.})}$$

- This test directly compares the best-fit suppressed model to the best fit pure power-law.
- Since $\mathcal{R}^2 \sim \chi_1^2$ we can estimate the (asymptotic) *Probability of False Acceptance*:

  $P_{\mathrm{FA}} \equiv$ probability of accepting the *suppressed* model given that the data are drawn from a *pure* power-law.

  **If** the data are drawn from a power-law **then** the chance that we would falsely accept either suppressed model is $P_{FA}$.

# Hague *Flux Suppression:* Summary

| Model | Name | Value | Stat | $^{+Sys}_{-Sys}$ | $p$-value | Conclusion |
|---|---|---|---|---|---|---|
| Power-Law | $\gamma$ | 2.78 | 0.02 | $^{-0.06}_{0.08}$ | $\geq 6\sigma$ (TP statistic) | Rejected |
| Double PL | $\gamma$ | 2.68 | 0.02 | $^{-0.05}_{0.08}$ | $\lg P_{FA} = -4.12$ | Favored |
| | $E_b$ | 35 | 2 | $^{7}_{7}$ | | |
| | $\delta$ | 4.22 | 0.22 | $^{-0.10}_{0.17}$ | | |
| Fermi PL | $\gamma$ | 2.63 | 0.02 | $^{-0.05}_{0.08}$ | $\lg P_{FA} = -4.29$ | Favored |
| | $E_{1/2}$ | 56 | $^{5}_{4}$ | $^{13}_{13}$ | | |
| | $w_c$ | 0.16 | $^{-0.03}_{0.02}$ | $^{-0.005}_{0.008}$ | | |

- The *preliminary result* is that we **can**:
  1. Reject the pure power-law model at a confidence level greater than six sigma.
  2. Favor either suppressed model with confidence better than $1/10,000$.
  3. Verify that the data are consistent with $E_{GZK} = 56 \pm 5$(stat)$\pm 15$(sys) EeV ... agrees with HiRes and with Berezinsky protons!
- This analysis alone **cannot** verify the GZK-cutoff, for that we need additional information on: CR composition (*e.g. all protons?*) and CR astrophysics (*e.g. sources uniformly distributed? constant source injection spectrum?*).

# Summary

- Auger has been a *different* experience ... and the (physical) challenge of a 55km $\times$ 55km detector at a remote, largely undeveloped site **cannot** be overstated!
- Certainly one of the high-points of working on Auger was our collaboration with the HiRes group!
- And at the craziest of times we have been able to have quite alot of fun developing “novel analyses”!

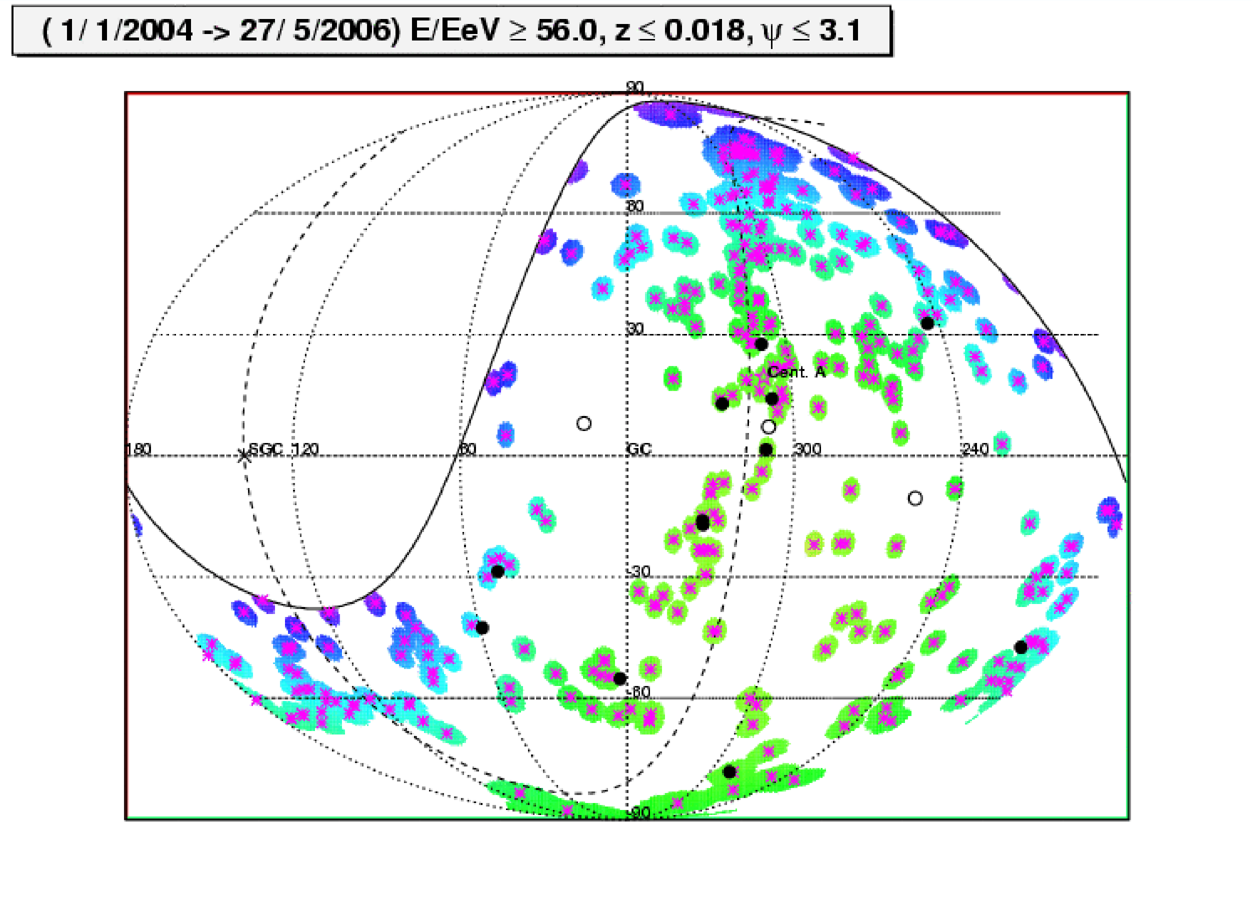
## Additional/backup slides

# Additional slides

# Which are *best* candidate sources?

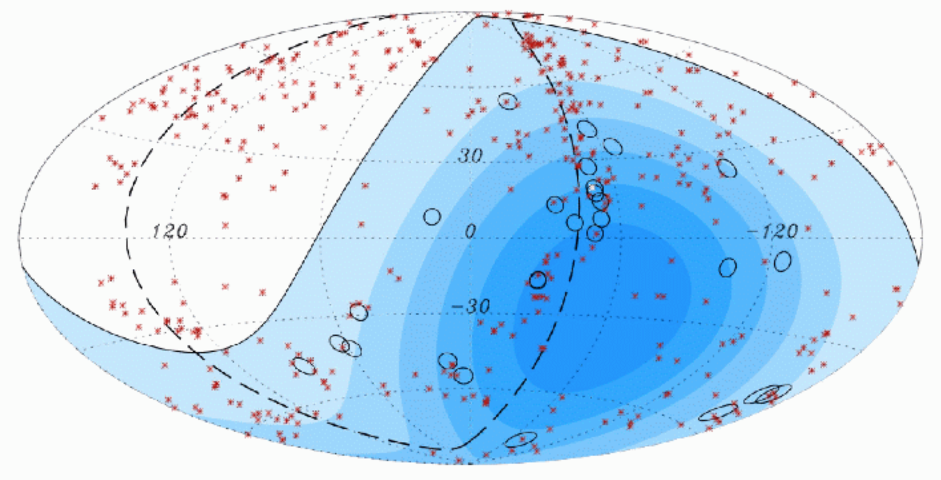
- Popular astrophysical sources for UHECRs include active galactic nuclei (AGNs) and gamma ray bursts (GRBs) ... **but no one knows: that is the Auger goal!**
- AGNs are super-massive black holes emitting jets of relativistic particles along the accretion disk rotation axis.
- Catalogs of AGNs provide a starting point ...
- So far the most significant correlations are with the $12^{th}$ Véron Cetty catalog
- **Centarus-A ($z = 0.0018$), shown above, is one of the nearby AGNs**

# Distribution of the 15 events above 56EeV




- Plot of nearby (Veron catalog) AGNs (*), each within a $3.1^\circ$ colored disk reflecting Auger acceptance, and CRs that correlate (filled circles) and that do not correlate (open circles).
- Miraculously, 12 of 15 CRs correlate ... *especially so* as the Véron catalog has a significant bias for galactic latitudes $|b| \lesssim 15^\circ$

# **So use a (1%)** *Running* **Prescription** on new data




- Depending on how you define *pass*, the *Running* Prescription passed in May (6/8) or in July (8/11) of 2007; the plot in *Science* includes events through Aug 31, 2007.
- *At a minimum, the Véron catalog: AGN maximum redshift and correlation angle, defines a limited area (effectively 21%) of the sky. Thus the Véron catalog AGN:CR correlation signal is evidence for a non-isotropic flux of CRs that is enhanced near known extra-galactic objects.*

# Alternate *Running* Prescription with Limited Error

- Brian Connolly (Segev BenZvi and Stefan Westerhoff) provided an alternative procedure ... that is of general interest! A generic version is now available as arXiv:0711.3937

- Philosophy: *All relevant information needed to infer parameters from an experiment is contained in the* ***observed*** *data.* **This is not true of the Auger Running prescription.**

- Motivation: Recall that the motivation for a running (*vs* fixed length) prescription is to be able to be as responsive as possible to data as they are collected!

- History: The technique comes from an "assembly line" defect analysis studied by Alexander Wald (1947). The relevant issue was how long to run a factory to ensure say $< 40\%$ of the cars were defective ... before shutting it down to re-tool the assembly lines. This technique was important enough to be classified by the U.S. government during W.W.II!

# Connolly/Wald *Running* Prescription (I)

Definitions, and values, for the case of our AGN:CR correlations:

- Background (random AGN:CR coincidence) probability: $p_0 = 0.21$
- Null hypothesis, $H_0$: corresponds to no signal, with correlation probability $p_0$
- Model *Signal* probability: $p_1$ (to be tested against $p_0$); this may be one value or a range of values: *e.g.* $p_1 > p_0$. For the (previous) *Running* prescription values: $p_1 = 0.57$ and $p_1 = 0.80$ were chosen.
- Model hypothesis, $H_1$: corresponds to a (model) signal, with correlation probability $p_1$
- The observed (*signal*) correlation probability in the new data: $p$
- Errors:
  - $H_0$ is true, but rejected by the test (*Type-I error*)
  - $H_0$ is false, but accepted by the test (*Type-II error*)
  - Limit probability of *Type-I* error: $\alpha = 0.01$
  - Limit probability of *Type-II* error: $\beta = 0.05$

# Connolly/Wald *Running* Prescription (II)

Sequential test of hypothesis $H_0$ *VS* $H_1$:

- Determine two positive constants: $A$ and $B$ (based on $\alpha$ and $\beta$ ... see below)
- After each new event calculate the probability ratio:

$$R = \frac{P(Data|H_1)}{P(Data|H_0)}$$

- If $R > A$ the running prescription is terminated with the rejection of $H_0$.
- If $R < B$ the running prescription is terminated with the acceptance of $H_0$.
- If $B < R < A$ the running prescription continues ... *i.e.* the result is inconclusive.
- Wald (1943) showed that: $A \geq \frac{1-\beta}{\alpha}$ and $B \leq \frac{\beta}{(1-\alpha)}$
- Furthermore Wald also showed that using "$=$" in the definitions for $A$ and $B$ provides protection against wrong decisions ... *i.e.* $\alpha$ and/or $\beta$ are not increased over the assigned values as long as they are $\lesssim 0.05$ ... **consistent with our choices**

# Connolly/Wald *Running* Prescription (III)

## Sequential test of Auger AGN:CR correlations:

- After each new event calculate the probability ratio:

$$R = \frac{p_1^k \cdot (1-p_1)^{n-k}}{p_0^k \cdot (1-p_0)^{n-k}}$$

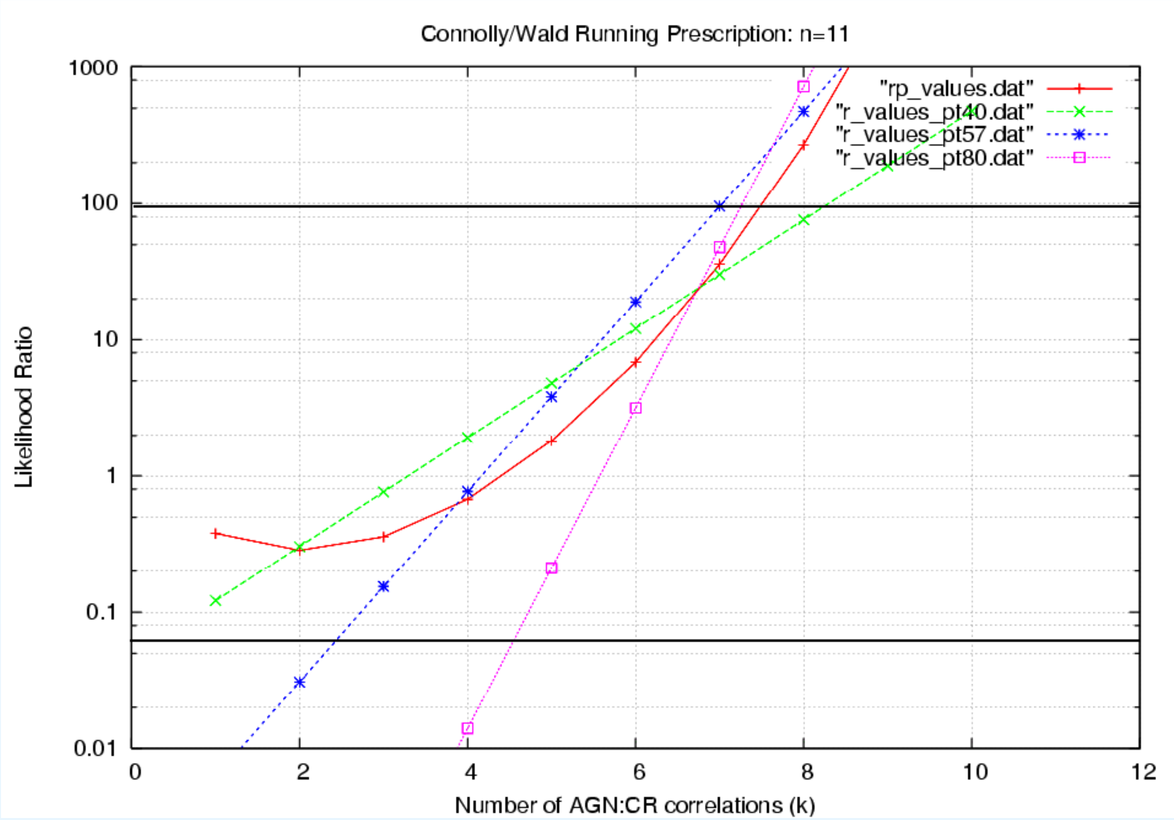
where $k$ events correlate out of $n$ total events and $p_0 = 0.21$. **But what value should we use for $p_1$?**

- One approach is to choose a model $p_1$ with $p_1 > p_0$ but less than, but possibly near, the correlation signal in the data, $p$; see arXiv:0711.3937.
- The new approach, proposed by Connolly, is to integrate over all possible values of $p_1$; then the ratio test becomes (for example):

$$R' = \frac{\int_0^1 p^k \cdot (1-p)^{n-k} dp}{p_0^k \cdot (1-p_0)^{n-k}} = \frac{B(k+1, n-k+1)}{p_0^k \cdot (1-p_0)^{n-k}}$$

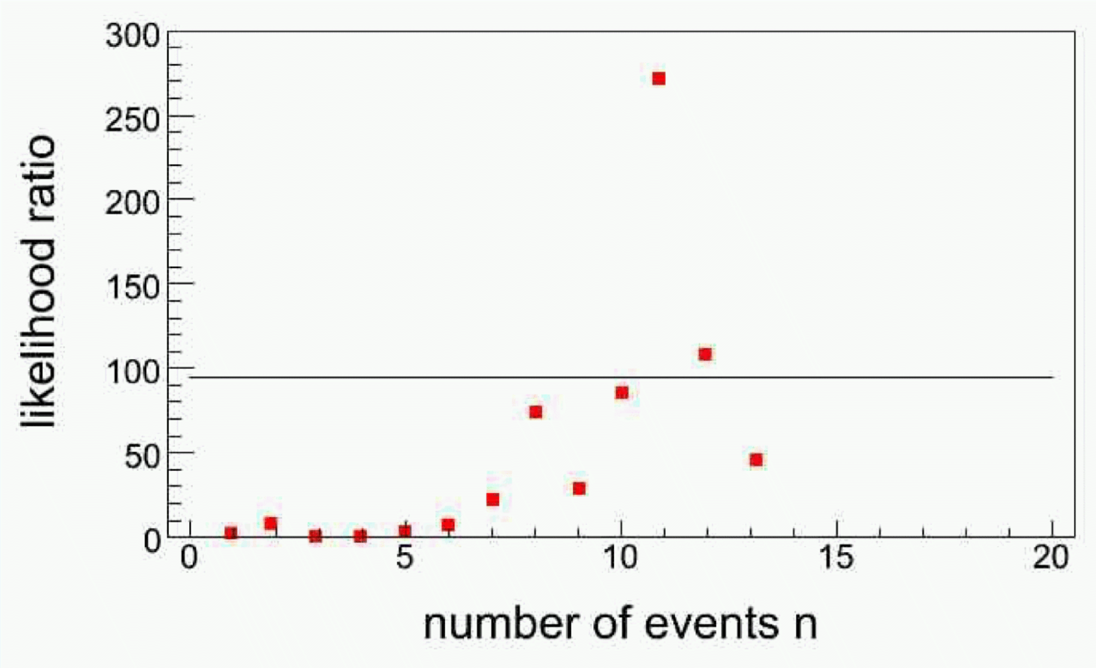
where B( ) is the beta function. This has now been validated in arXiv:0711.3937.

# Connolly/Wald *Running* Prescription (IV)




- With our choice of $\alpha = 0.01$ and $\beta = 0.05$ then $A = 95$ and $B = 0.0505$
- For a data sample ($n$) of 11 events, how sensitive are $R'$, and/or $R$, to the observed number of correlations ($k$)?
- *Plot*: shows $R'$ and $R$ (for three values of $p_1$: 0.4, 0.57, 0.8) *vs* $k$

# Connolly/Wald *Running* Prescription (V)




- With our choice of $\alpha = 0.01$ and $\beta = 0.05$ then $A = 95$ and $B = 0.0505$
- If $R' < 0.0505$ the null hypothesis is accepted ... *i.e. this is* **evidence against a signal**
- If $0.0505 < R' < 95$ ... keep going *i.e. we simply do not know*!
- If $R' > 95$ the null hypothesis is rejected ... *i.e. this is* **evidence for a signal ...**

  This occurred when $k = 8$ correlations were observed in $n = 11$ total events