A Three-point (Shape Strength) Cosmic Ray Anisotropy Search Method

John A.J. Matthews

johnm@phys.unm.edu

University of New Mexico

Albuquerque, NM 87131

FOR

Bernie Becker, Michael Gold, Doug Hague, and the Pierre Auger Collaboration

Anisotropy searches: *Alan Watson* *(Auger Celebration, 2005)*

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

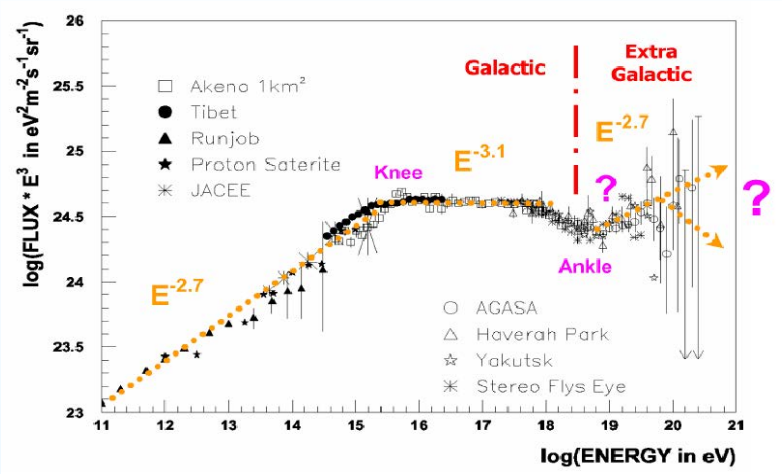
Magnetic Fields are the problem:

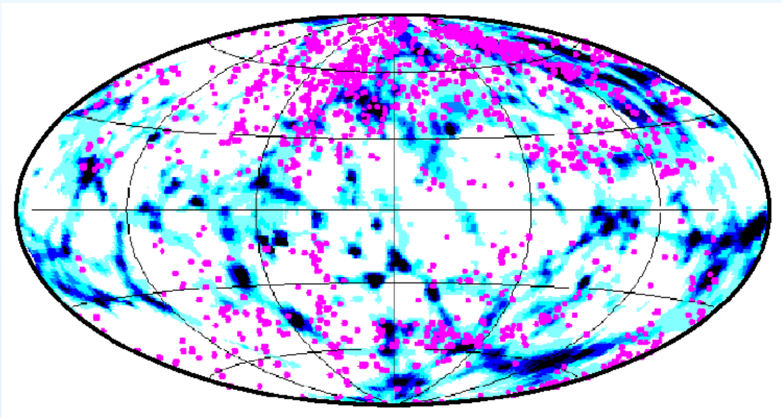
While gamma-rays and neutrinos are 'blind' to magnetic fields, cosmic rays are charged particles, the nuclei of atoms.

Like the drunken man's walk!

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

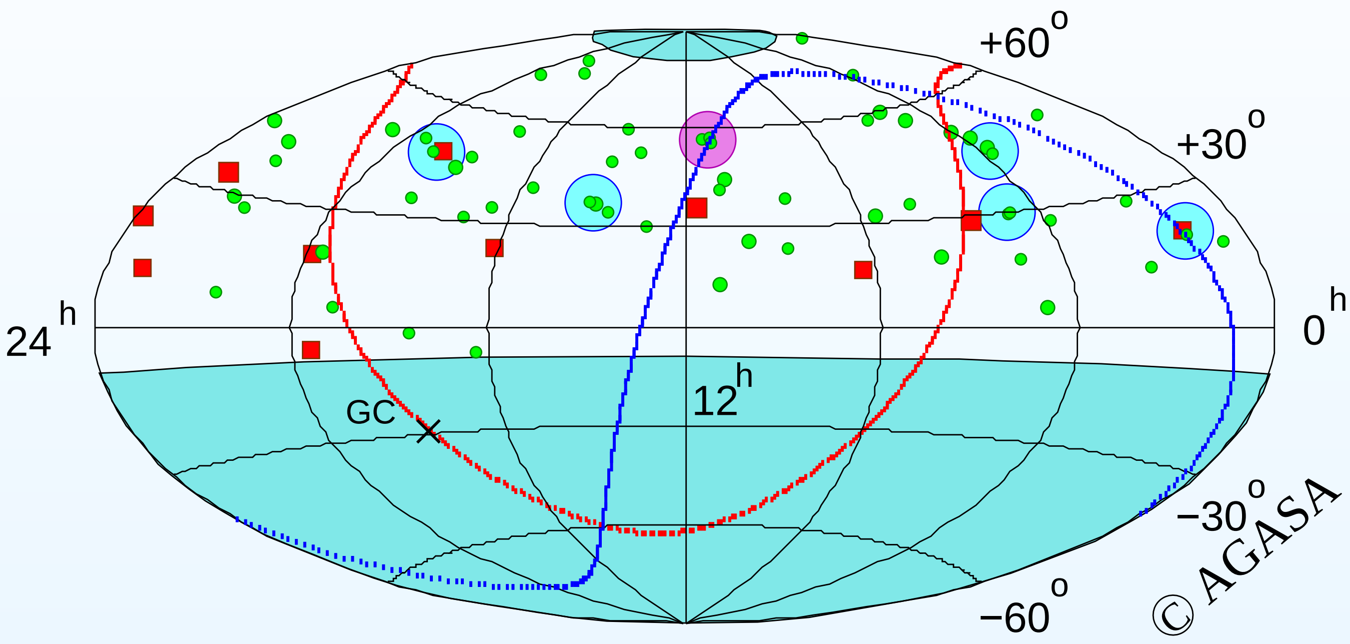
Motivation for cosmic ray anisotropy:





- 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*
- HiRes stereo, with > 3-times the exposure, has **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

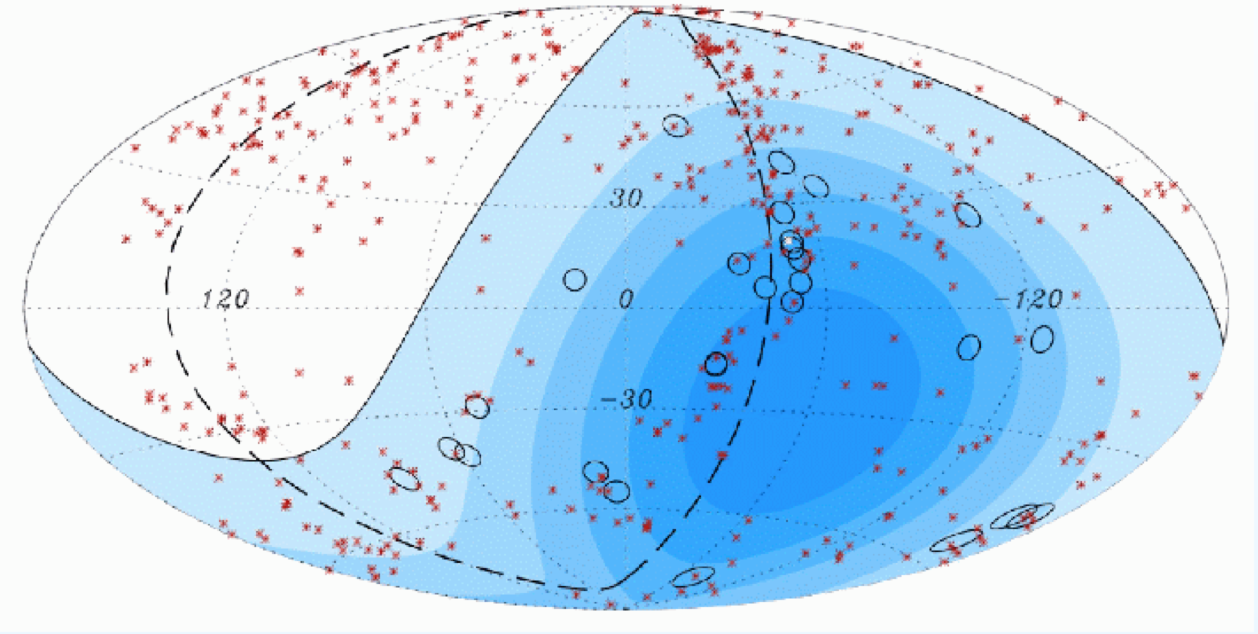
D. S. Gorbunov[a 1)], *P. G. Tinyakov*[b,a], *I. I. Tkachev*[c,a], *S. V. Troitsky*[a]

[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

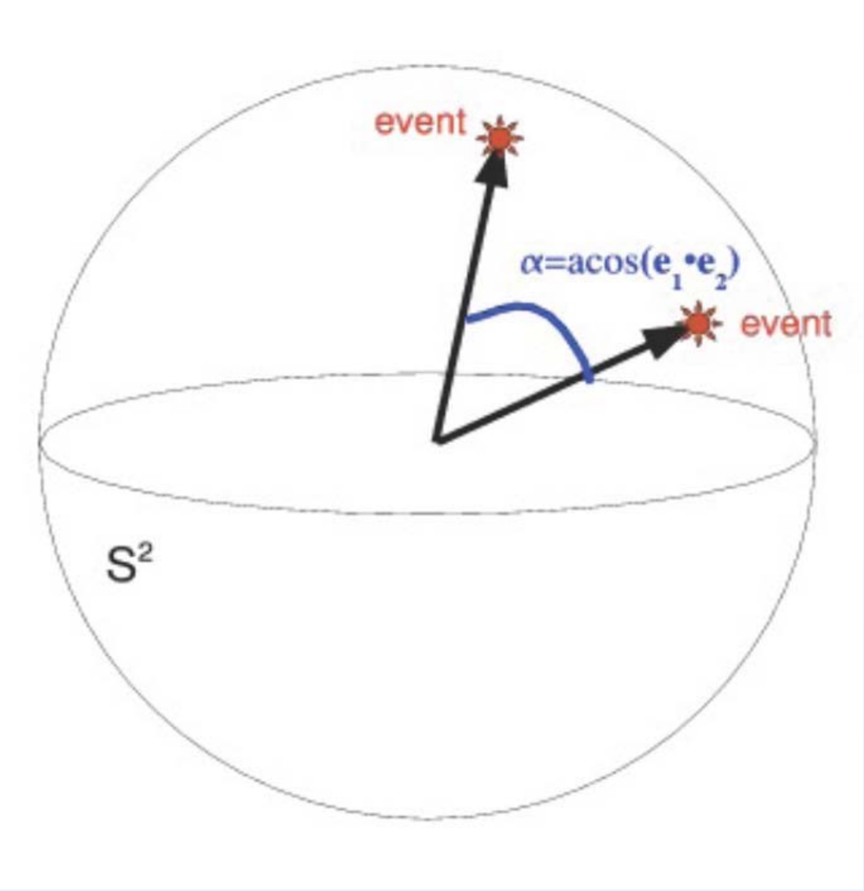


- The most compelling observational evidence consistent with astrophysical expectations of anisotropy is arguably the 27 events with energy > 57 EeV observed by Auger.
- At a minimum, the Véron catalog: AGN maximum redshift and correlation angle, defines a limited area (effectively 21%) of the sky. Thus 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:

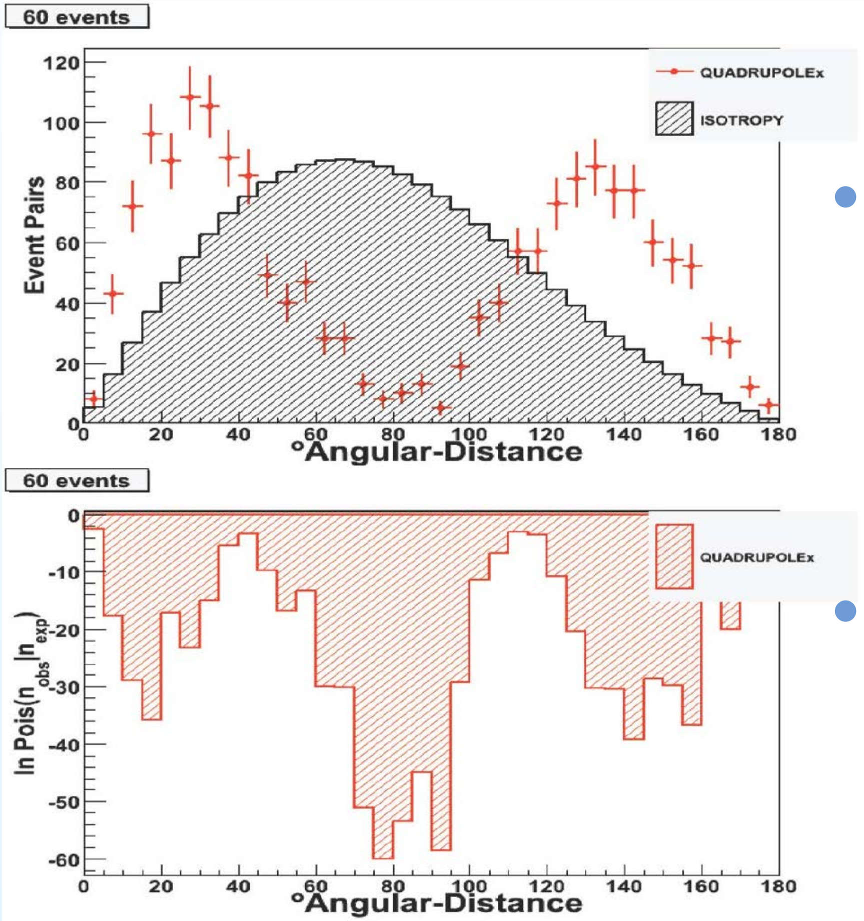
- Catalog dependent (cross-correlation) studies are not without issues: *e.g.* penalty factors for scans over different catalogs, issues related to *brightness limited* catalogs, and/or the need to restrict the data to match the limited sky coverage of individual catalogs.
- However with limited statistics, 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 on the LANL archive: http://xxx.lanl.gov/abs/0905.4488

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 with 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”!

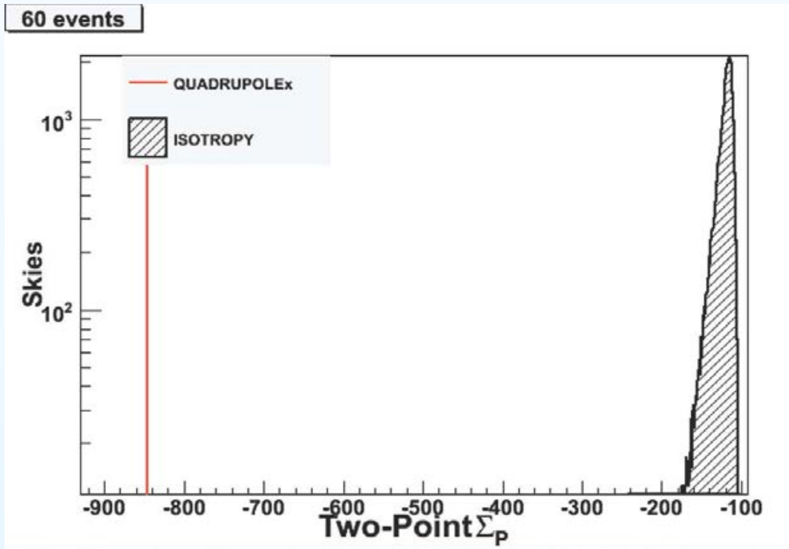
A 2-point method (*toy example - I*):



- *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:

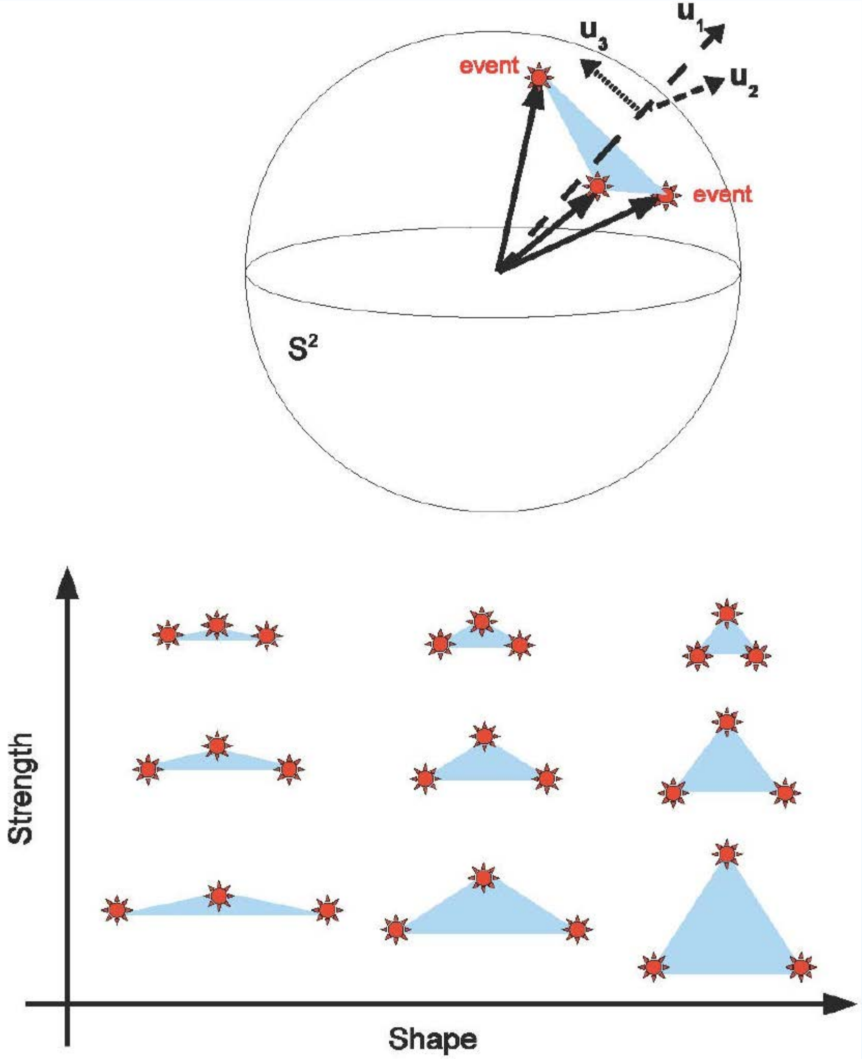
  $$\Sigma_P = \Sigma_{i=1}^{N_{bins}} \ln \mathrm{P}_i(n_{obs}|n_{exp})$$

  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:

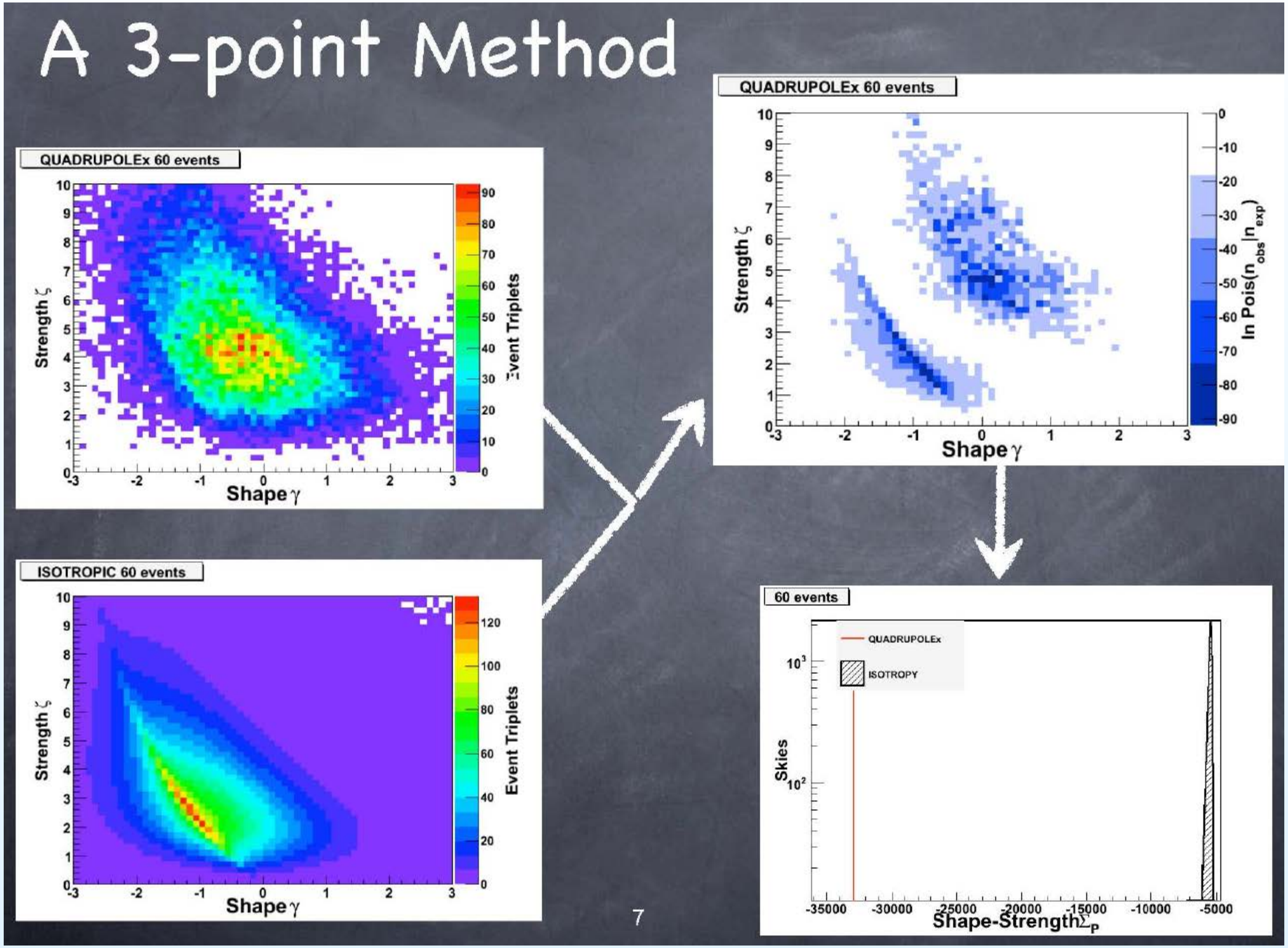
  Shape: $\gamma = log(\ \frac{log(\tau_1/\tau_2)}{log(\tau_2/\tau_3)}\ )$

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

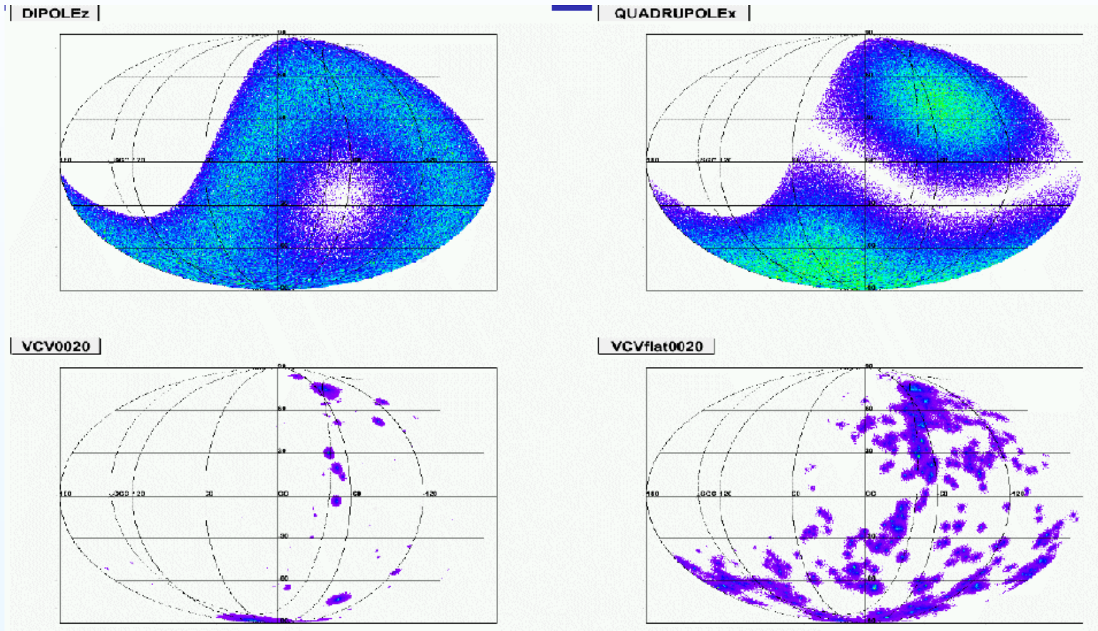
  Strength: $\zeta = log(\ \tau_1/\tau_3\ )$

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

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

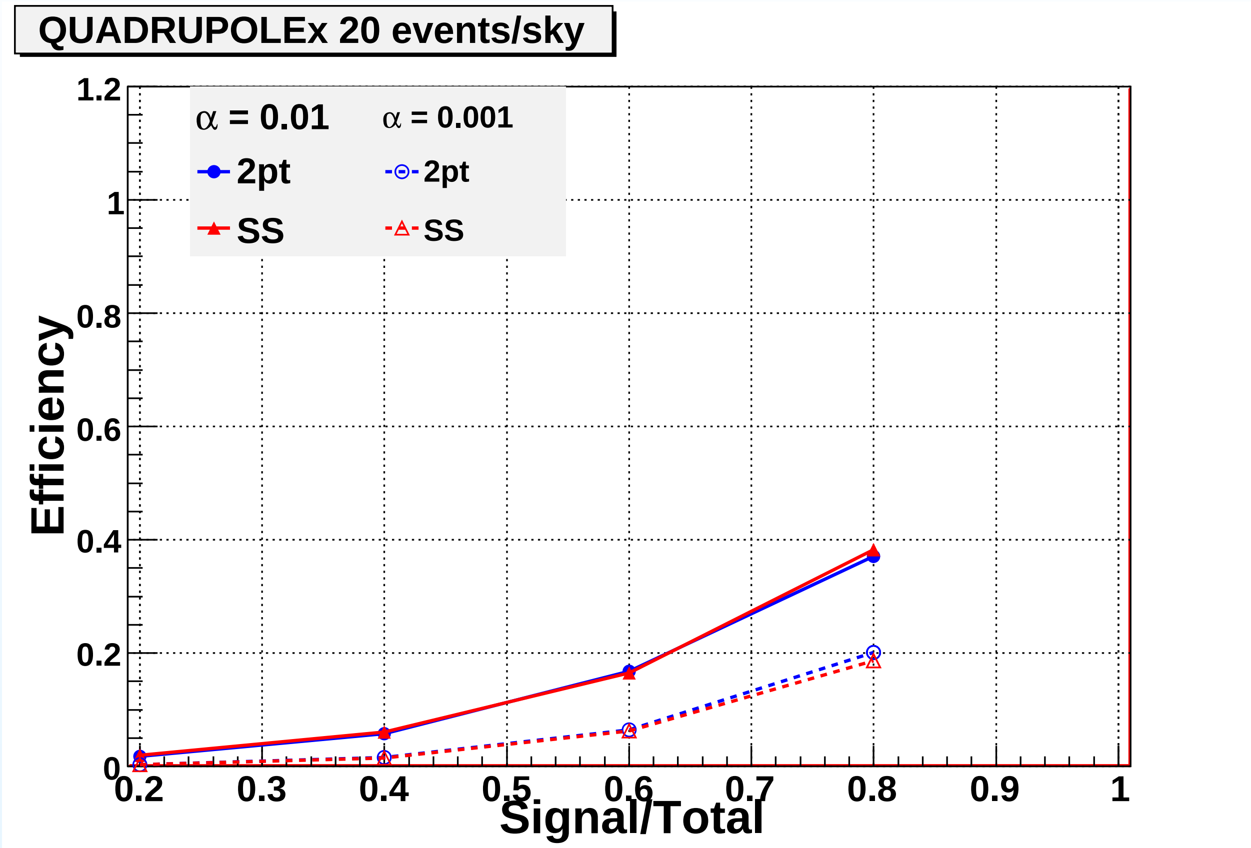


Catalog independent analyses (I)



- 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.

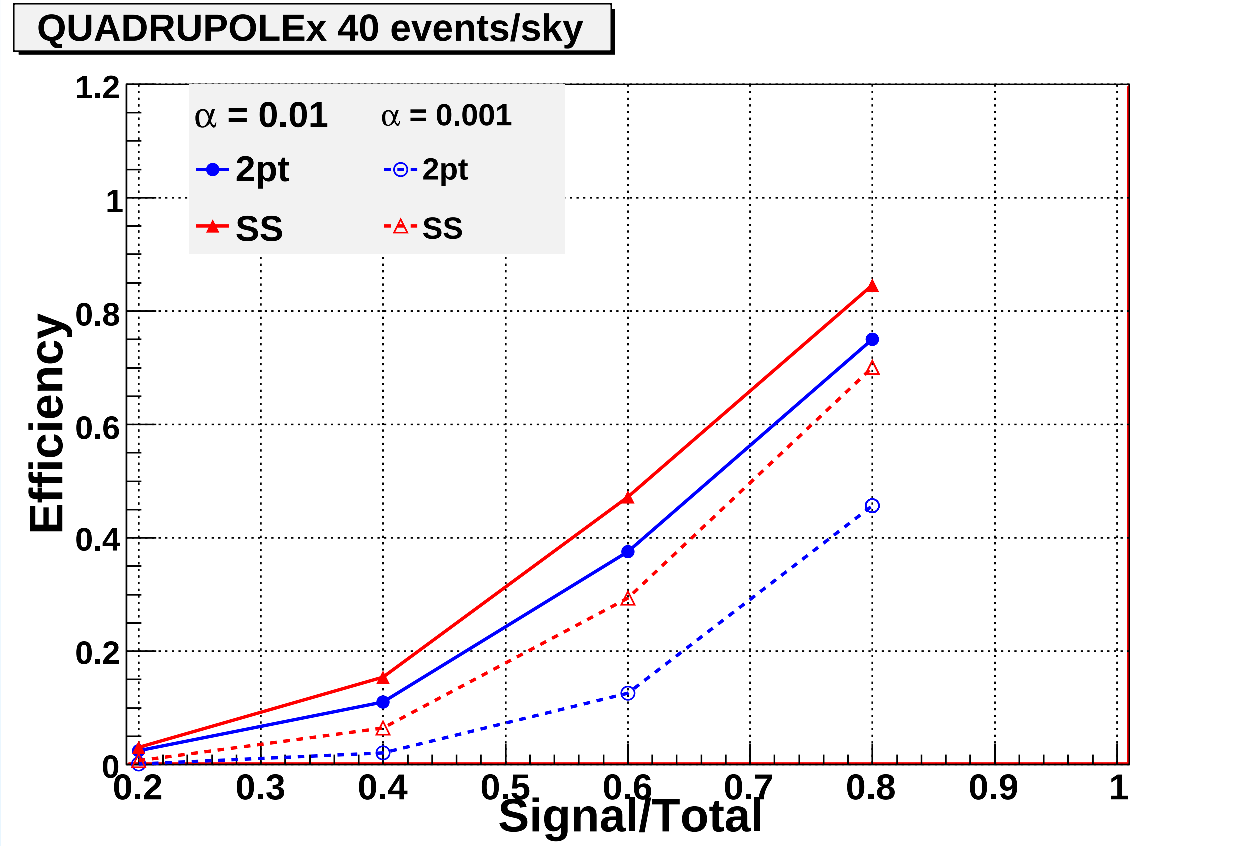
Catalog independent analyses (II)



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

20 events

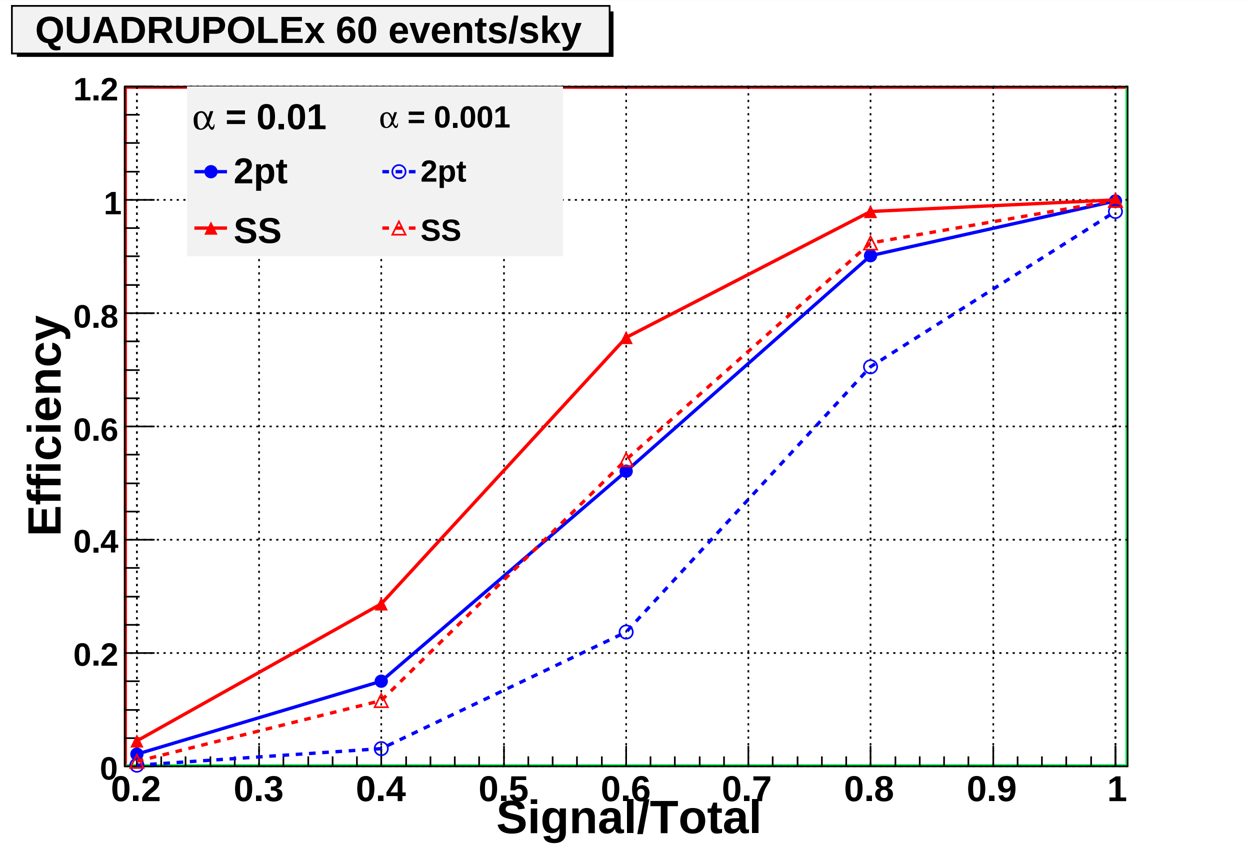
Catalog independent analyses (III)



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

40 events

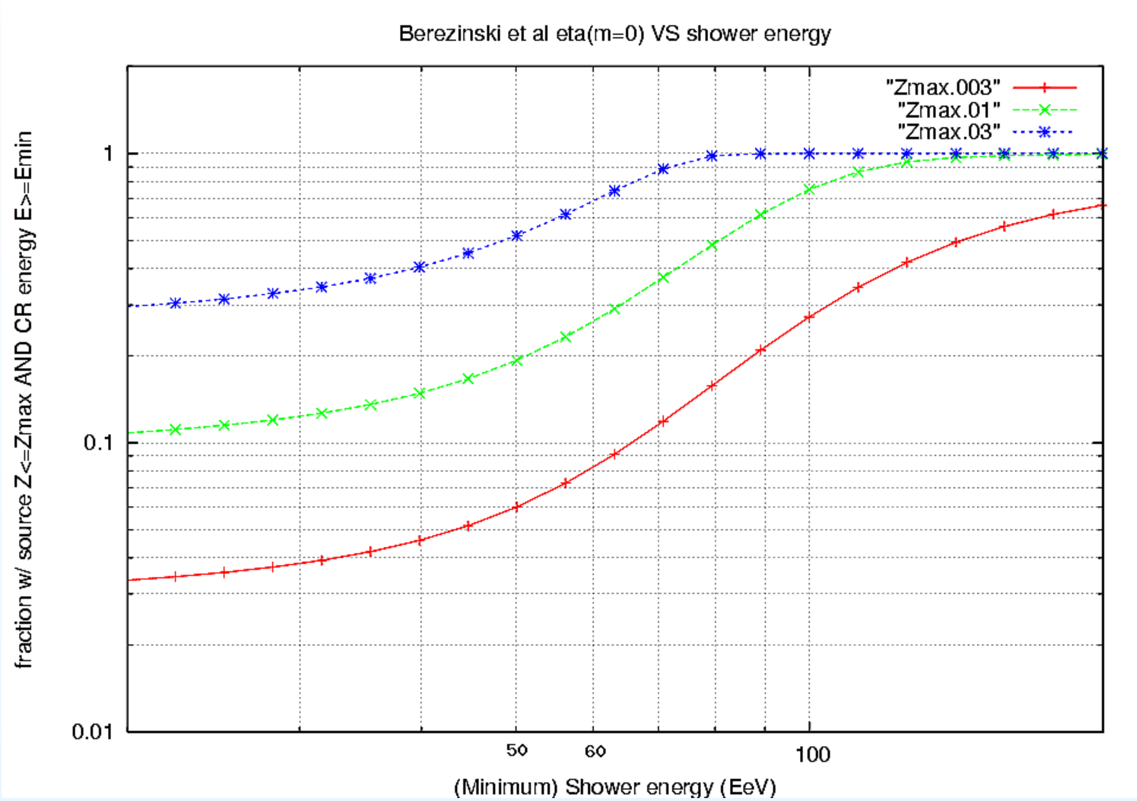
Catalog independent analyses (IV)



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

60 events

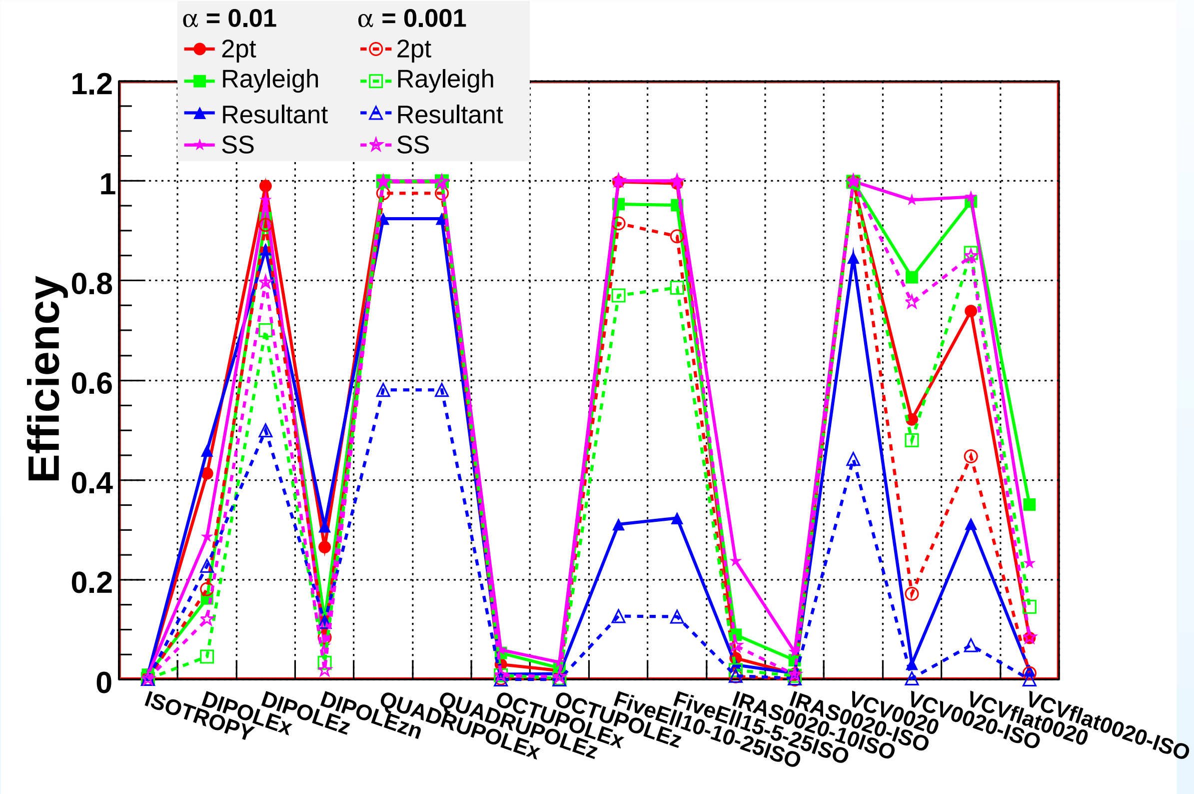
Catalog independent analyses (V-a)



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)

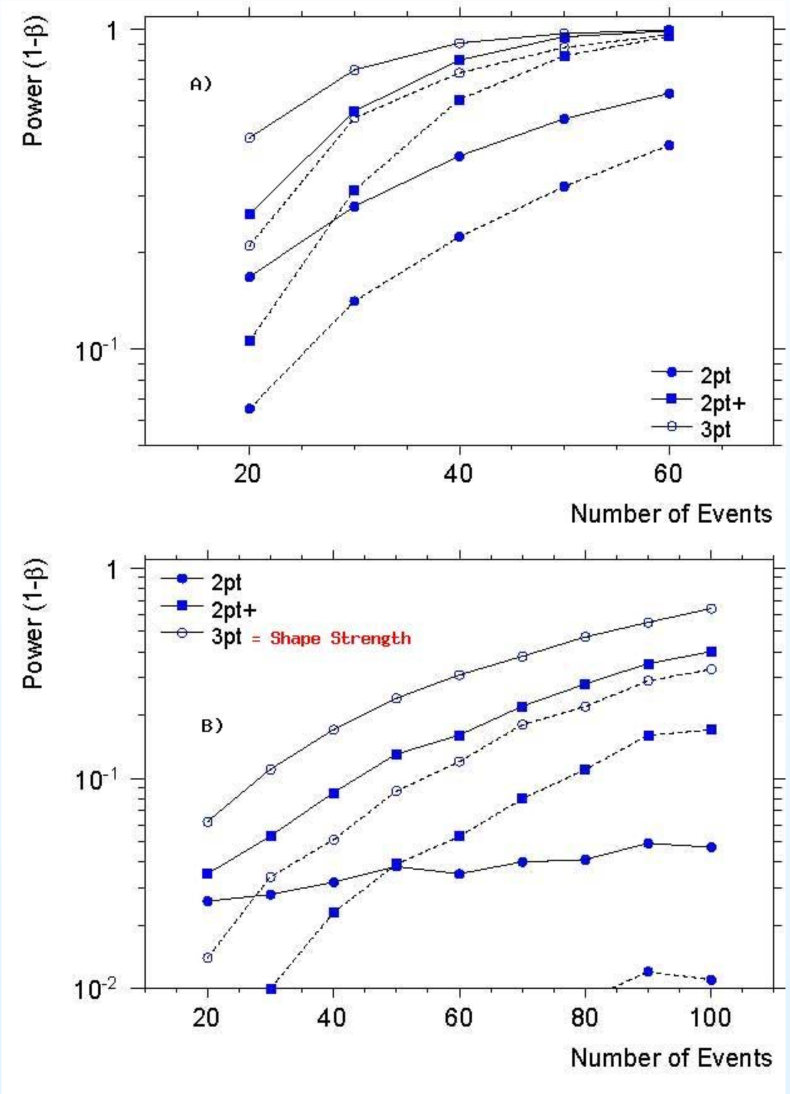


Based on simulated samples (*ie mock data*) from hypothetical sources, we find:

- 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.

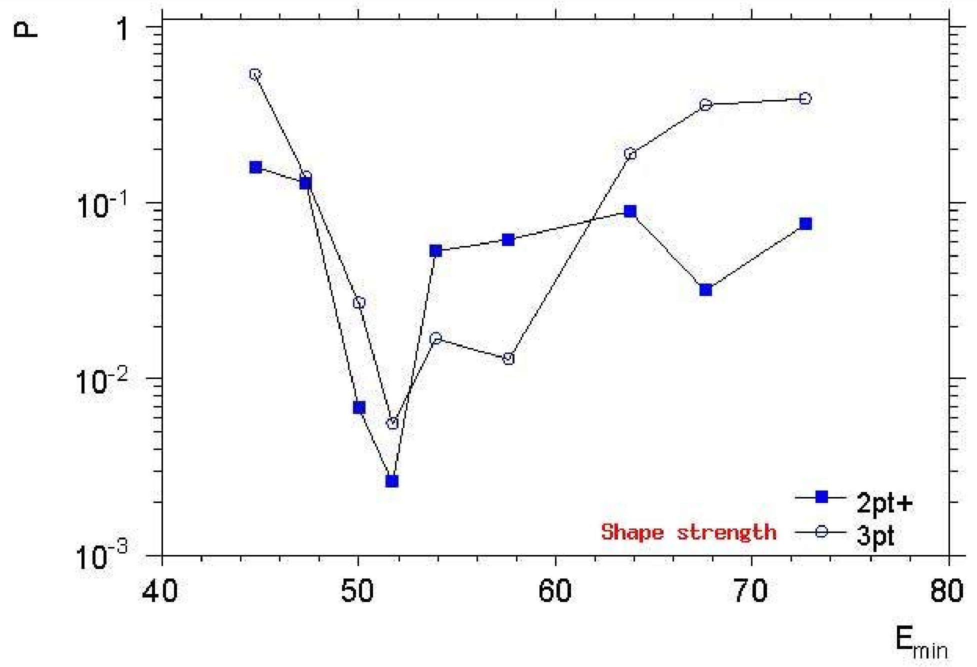
C.I.A. application to Auger data (I)

The University of New Mexico



- 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)

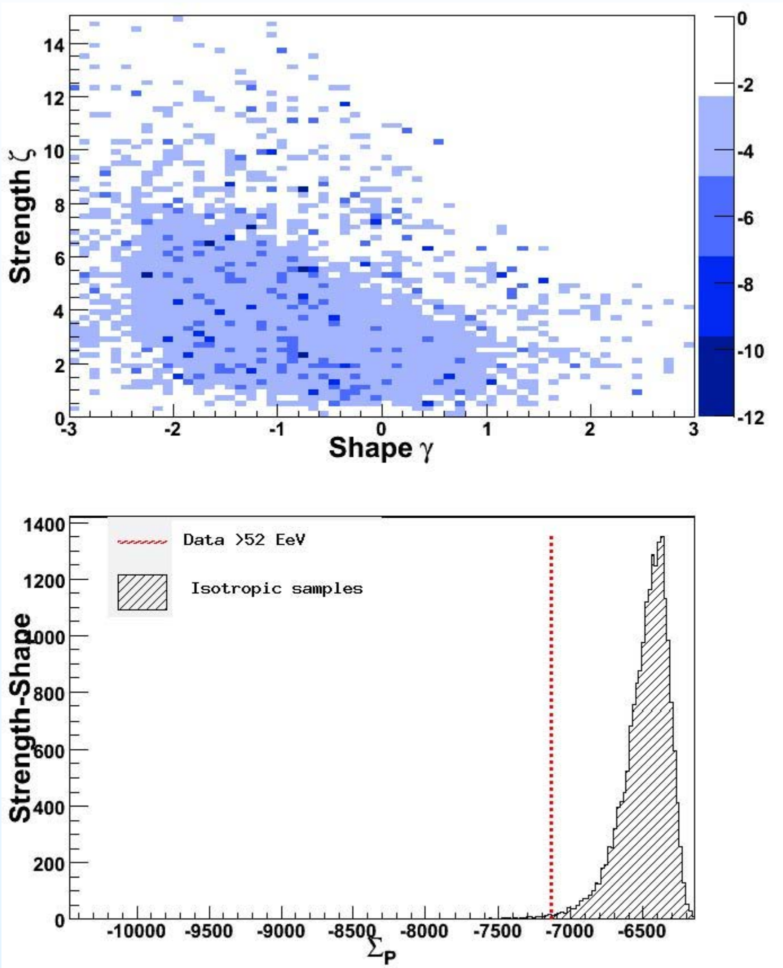


Scan of Auger data: January 1, 2004 to March 31, 2009 starting with the 20 highest energy events then in steps of 10 events to 100 events.

- 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)

The University of New Mexico



- 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 ~ 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