Guilt by Association: Finding Cosmic Ray Sources

David Ruppert

Operations Research & Information Engineering
and Department of Statistical Science, Cornell University

Jun 24, 2011

Collaborators

- Kunlaya Soiaporn, Graduate Student, ORIE
- Tom Loredo, Research Associate, Astronomy
- Dave Chernoff, Professor, Astronomy
- Ira Wasserman, Professor and Chair, Astronomy

Outline

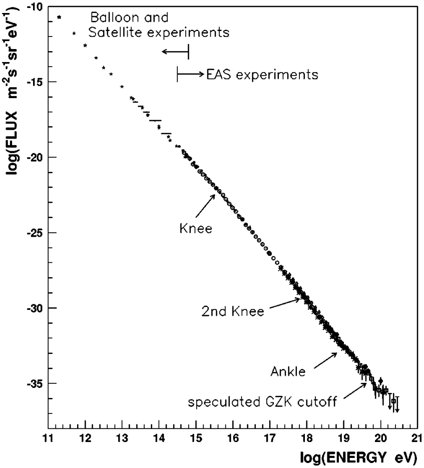
- Cosmic Rays
    - this research is about ultra-high energy cosmic rays
- Active galactic nuclei (AGNs)
    - a prime suspect as the source of ultra-high energy cosmic rays
- Association models
    - associate cosmic rays with AGNs
    - can we "convict" AGNs as the source

Caveat: **This project is still "work in progress"**

What Are Cosmic Rays?

- Cosmic rays are atomic nuclei
- First detected in 1912 by Victor Hess who ascended in a balloon to 5 km
- Range in energy from 10^7 to 10^{20} eV
  - eV $=$ electron volt
- Spectrum is a power law $F \propto E^{-\alpha}$
  - $F =$ flux
  - $E =$ energy
- Detailed look at F versus E (log-log plot) suggests several sources

Cosmic Ray Spectrum



Where Do Cosmic Rays Originate?

- Cosmic rays are charged particles
  - therefore they are deflected by magnetic fields
  - so it is not obvious where they originate
- Sources of cosmic rays could be
  - supernovae
  - pulsars
  - stars with strong winds
  - black holes
- Active galactic nuclei (AGNs) are a prime suspect of cosmic rays at highest energies
  - only AGNs seem capable of accelerating particles to such high energies

Ultra-High Energy Cosmic Rays (UHECRs)

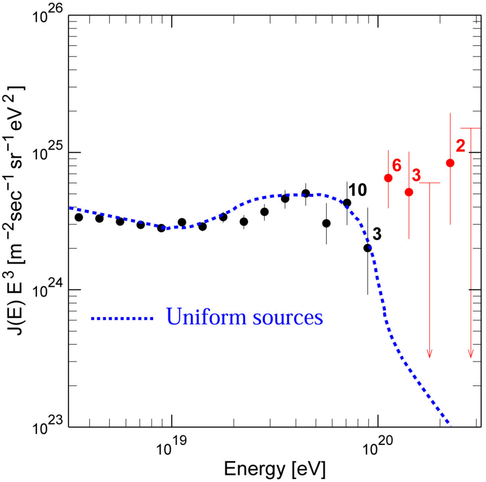
- Our research focuses on cosmic rays of highest energies
- Cosmic ray with E $> 10^{20}$ eV observed in 1962
- 1991: particle with E $\approx 3 \times 10^{20}$ observed
  - same kinetic energy as a baseball at 60 mph
  - over 10 million times more energy than most energetic particles at Large Hadron Collider

- Not confided to galaxy of origin
- Interact with cosmic microwave background
  - called the GZK cutoff
  - So UHERCs must come from within approximately 100 megaparsecs (Mpc)
  - 1 parsec $\approx$ 3.26 light-years
- Closer galaxies are more likely sources
- Ultra-high energy cosmic rays create giant air showers of particles
  - first discovered by Pierre Auger (1899–1993)

Early Cosmic Ray Detectors: AGASA

- Akeno Giant Air Shower Array (AGASA) is a very large surface array in Japan
- In operation February 1990 – January 2004
- Covers an area of 100 km^2 and consists of 111 surface detectors and 27 muon detectors

Agasa Spectrum: No GZK Cutoff!!!



Early Cosmic Ray Detectors: Hi-Res

- High Resolution Fly's Eye or HiRes detector observatory
- Operated in the western Utah desert
  - from May 1997 until April 2006
- Utilized the atmospheric fluorescence technique
- Made the first observation of the GZK cutoff
  - So conflicts with AGASA findings

Typical Fluxes

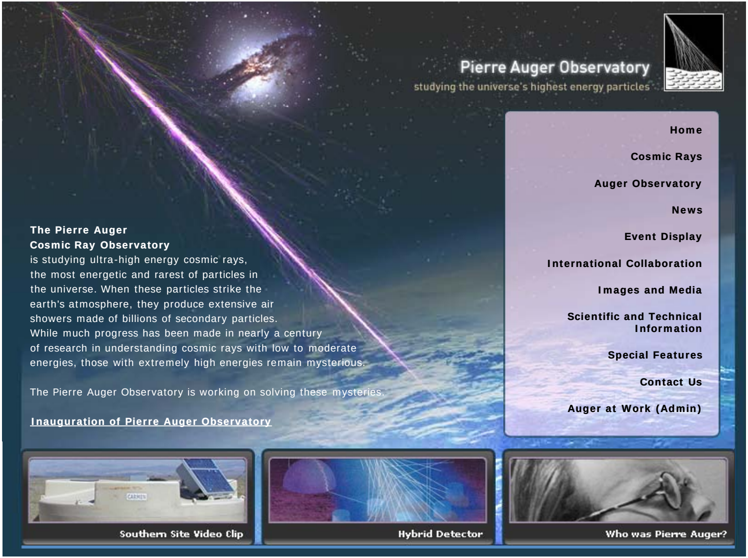
- Fluxes vary by a factor of 10^{32} from one end of the spectrum to the other
  - At low end of spectrum: 1 particle m^{-2} s^{-1}
  - At high end: 1 particle km^{-2} century^{-1}
- The Pierre Auger Observatory can detect cosmic rays at the high end
  - covers 3000 km^2

Pierre Auger Observatory:

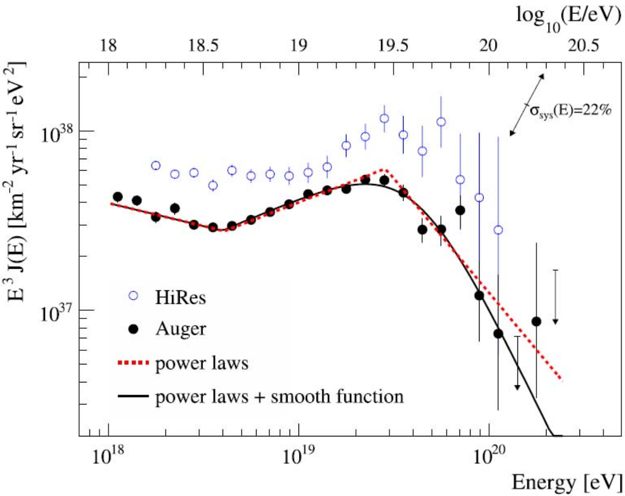
- Largest and most sensitive cosmic ray detector to date
- In Argentina
- Uses air fluorescence telescopes and surface detectors
- Operations began in 2008
- Has detected about 70 UHECRs

Pierre Auger Home Page

Pierre Auger Observatory

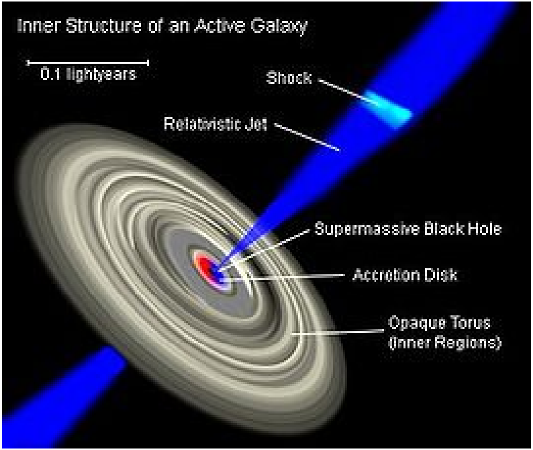


High-energy Cosmic Ray Spectrum



What is an Active Galactic Nucleus (AGN)?

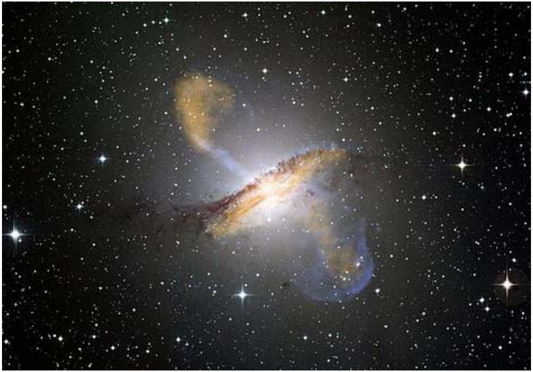
- An AGN is a compact region at the center of a galaxy with high electromagnetic luminosity
  - Example: Quasar
- Activity is believed to come from the accretion of mass by a supermassive black hole
- Our galaxy also harbors a supermassive black hole
  - but the Milky Way does not seem to be active at present



Source: Wikipedia

Radio Galaxy Centaurus A (NGC 5128) – Visible Spectrum

Source: Wikipedia



870-micron submillimeter = orange; X-ray = blue; visible light = close to true color

Source: Wikipedia

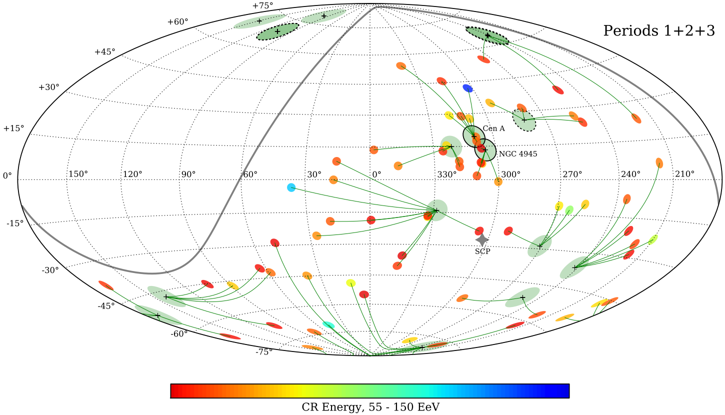
- We used all AGNs within 15 Mpc (megaparsecs)
- To decide which galaxies were AGNs astronomers did the following:
    - Start with 64 infrared-bright galaxies within 15 Mpc
    - Select AGNs based on an infrared spectral line of neon
    - It takes so much energy to excite this line that AGN activity is the only likely cause
    - This line was seen in 17 of the 64 IR-bright galaxies

- PAO has detected 69 UHECRs with energy $\geq 5.5 \times 10^{19}$ eV

Period	Dates	Exposure (km^2 sr y)	No. of UHECRs detected
1	01-01-04 – 05-26-06	4390	14
2	05-27-06 – 08-31-07	4500	13
3	09-01-07 – 12-31-09	11480	42

- The CR flux from all 3 periods is $(14 + 13 + 42)/(4\pi \times \text{Total Exposure}) = 0.043\ \text{km}^{-2}\text{yr}^{-1}$
- According to GZK limit, the CRs with energies $\gtrsim 5 \times 10^{19}$ should interact with cosmic microwave background photons, and should almost never reach the earth from distances larger than 50 Mpc

UHECR – AGN Association: Evidence From First 69 CRs



Energy ranges from 55 EeV to 142 EeV: 1 EeV $= 10^{18}$ eV

Goals of Our Research

1. Compare models with different source populations
   - including a "null" or isotropic source
2. Estimate the amount of scattering by cosmic magnetic fields
3. Ascertain which cosmic rays are associated with specific sources (with high probability)

Goals of Our Research, Cont.

4. Estimate flux of each source of cosmic rays
5. Estimate luminosity function parameters
6. Investigate whether cosmic rays from a source are scattered independently ("buckshot model") or undergo nearly identical scattered ("radiant model")

Our model has 4 levels:

1. **candidate source population (e.g., AGNs)**
   - distribution for source luminosities (a “luminosity function”)
   - “zeroth” source = an isotropic background component with uncertain luminosity.

Null model: All observed cosmic rays are from the zeroth source

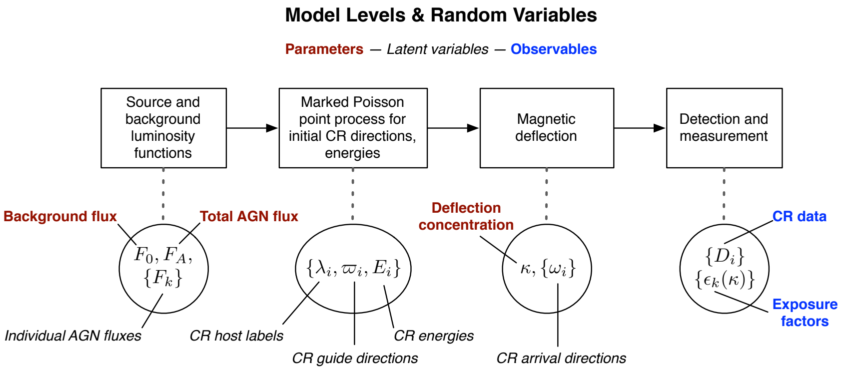
2. **marked Poisson point process model for latent cosmic ray properties**
   - the arrival times have a homogeneous intensity measure in time
   - the marks include
     - latent "guide" directions for the cosmic rays
     - the cosmic ray energies
     - latent categorical labels identifying the source of each ray

3. Model for magnetic deflection of the rays, scattering their directions from the guide directions

Multilevel Modeling: Level 4

4. Measurement model with directional uncertainties and accounting for truncation and thinning

Four Levels and Associated Parameters



- We consider the 17 active galactic nuclei (AGNs) in the volume-complete catalog of Goulding (2010) as candidate sources. The catalog is complete to 15 Mpc.
- An isotropic background is included as a“zeroth” source
- 3 different models:
  - M_0 : only isotropic background source
  - M_1: isotropic background source + 17 AGNs
  - M_2: isotropic background source + 2 AGNs: Centaurus A (NGC 5128) and NGC 4945, which are the two closest AGNs

Models, cont.

- CR arrivals follow a time-homogeneous Poisson process with rate depending on the fluxes and exposure factors of sources
- The measurement error of CR direction is modeled using Fisher distribution with concentration parameter corresponding to angular uncertainty of 0.9°

- The magnetic deflection of each CR direction is modeled using a Fisher distribution with concentration parameter κ ($\kappa \approx \frac{2.3}{\sigma^2}$ for 2-d Gaussian approximation with standard deviation σ radians)
- We treat κ as an unknown parameter

Bayesian Hierarchical Model

- We use a 4-level hierarchical model, schematically shown above
- $F_0 \sim$ exponential(scale=s),

  $F_A \sim$ exponential(scale=s),

  $F_k = w_k F_A$, where $w_k \propto 1/$squared distance to AGN$_k$,

  $\sum_{k \geq 1} w_k = 1$

  $f = \frac{F_A}{F_A + F_0}$

- $\Pr\{\lambda_i = k | F_0, F_A\} \propto F_k \epsilon_k$
- $P(\text{Data}|F_0, F_A, \lambda) = e^{-\sum F_k \epsilon_k} \left(\sum F_k \epsilon_k\right)^{N_C} \prod_i \frac{f_{\lambda_i,i}}{\epsilon_{\lambda_i}}$
- $f_{\lambda_i,i}$ is the marginal likelihood attributing CR$_i$ to AGN λ_i,

taking into account

  - the measurement error,
  - the exposure toward the AGN and
  - the magnetic deflection,

- N_C is the number of CRs

Prior Specification

- For M_0, F_0 has exponential prior with scale $2s$
- For M_1 and M_2, both F_0 and F_A have exponential prior with scale s
- In every model, the expected total fluxes are the same apriori. We choose $s \approx 0.063$ km^{-2}yr^{-1}, based on the data from the two previously operated observatories, AGASA and HiRes

Markov Chain Monte Carlo – Initialization

Initialize:

- $F_0 \sim$ exponential(scale$=s$),
- $F_A \sim$ exponential(scale$=s$),
- $F_k = w_k F_A,\ k = 1, 2, \ldots, M$,
- $\Pr\{\lambda_i = k | F_0, F_A\} \propto F_k \epsilon_k$

Markov Chain Monte Carlo – Iteration

- Gibbs sampling:

$$
\begin{aligned}
\mathsf{P}(F_A|F_0,\lambda,\text{Data}) & \\
&\sim \text{gamma}\left(1+\sum_{k\geq 1} m_k(\lambda), \frac{1}{\frac{1}{s}+\sum_{k\geq 1} w_k\epsilon_k}\right) \\
\mathsf{P}(F_0|F_A,\lambda,\text{Data}) &\sim \text{gamma}\left(1+m_0(\lambda), \frac{1}{\frac{1}{s}+\epsilon_0}\right) \\
\mathsf{P}(\lambda_i=k|F_A,F_0,\text{Data}) &\propto f_{k,i}F_k
\end{aligned}
$$

- $m_k(\lambda)$ is the number of CRs assigned to source k according to λ

Marginal Likelihood – Exact

- Marginal likelihoods are available in closed form in all models, but require summing over all possible values of λ for M_1, M_2

Marginal Likelihood – Chib's Method

- Chib's estimate for the marginal likelihood is used for models $m = 1, 2$:

$$\ell_m = \frac{\mathsf{P}\left(\text{Data}|F_0^*, F_A^*, \lambda^*\right)\mathsf{P}\left(F_0^*\right)\mathsf{P}\left(F_A^*\right)\mathsf{P}\left(\lambda^*|F_0^*, F_A^*\right)}{\mathsf{P}\left(F_0^*, F_A^*, \lambda^*|\text{Data}\right)}$$

where the denominator can be expressed as

$$\mathsf{P}(F_A^*|F_0^*, \lambda^*, \text{Data})\mathsf{P}(F_0^*|\lambda^*, \text{Data})\mathsf{P}(\lambda^*|\text{Data})$$

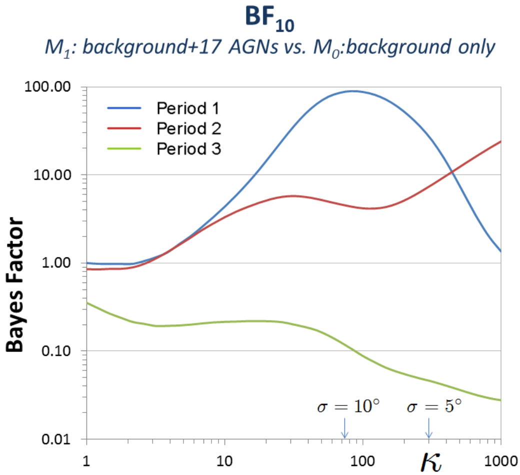
$$= \mathsf{P}(F_A^*|\lambda^*)\mathsf{P}(F_0^*|\lambda^*)\mathsf{P}(\lambda^*|\text{Data})$$

- F_0^*, F_A^*, λ^* are chosen from high-posterior points
- $\mathsf{P}(\lambda^*|\text{Data})$ is estimated using Gibbs sampling. All other terms are computed analytically.

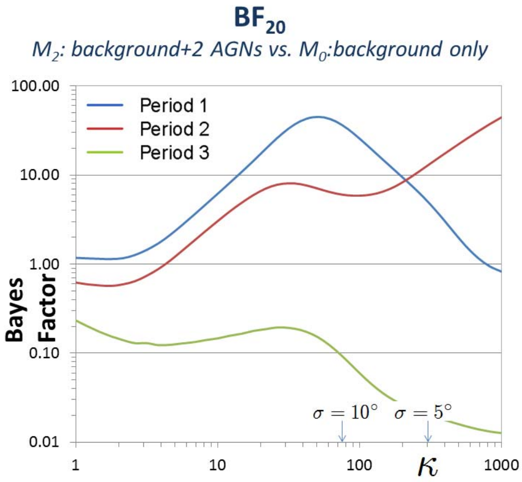
- We compare models 1,2 to model 0. The Bayes factors are computed as

$$\text{BF}_{10} = \frac{\ell_1}{\ell_0}\ ,\ \text{BF}_{20} = \frac{\ell_2}{\ell_0}$$

Bayes Factor Plot – 17 AGNs



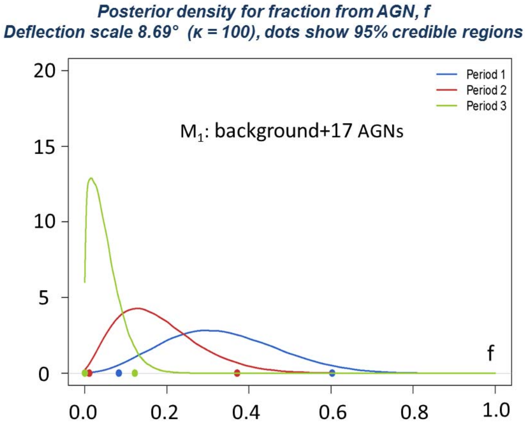
Bayes Factor Plot – 2 AGNs



Overall Bayes Factors for log-flat prior over [1,1000]

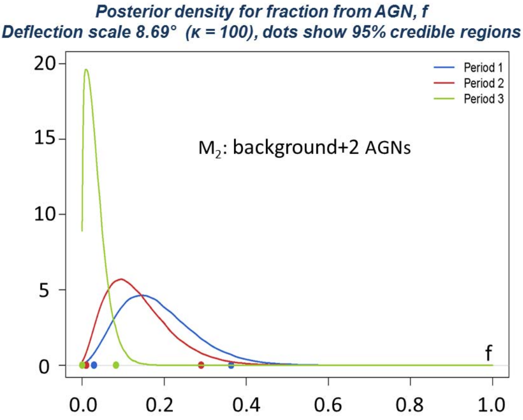
	Period 1	Period 2	Period 3
BF_{10}	25.27	5.60	0.15
BF_{20}	11.99	8.67	0.11

Posterior density of f, Model M1



$$f := F_A/(F_A + F_0)$$

Posterior density of f, Model M2



$$f := F_A/(F_A + F_0)$$

Summary

- The strength of the evidence for association with these AGNs differs markedly from period to period
  - we will investigate if time inhomogeneity can be attributed to random variation
  - if not, then there is a contradiction
  - magnetic scattering implies that any time heterogeneity would be on the order of thousands of years
- Presuming these AGNs are CR sources, $\sim 10\%$ of PAO CRs may come from them, but a significant fraction appears to originate elsewhere

What Next?

- Consider other CR luminosity functions
- Investigate the significance of period-to-period variations and either
  - Develop a changepoint model (if significant)
  - Aggregate the three periods (if not significant)
- Compare models with different source populations
- Consider different magnetic deflection models
- Wait for more data