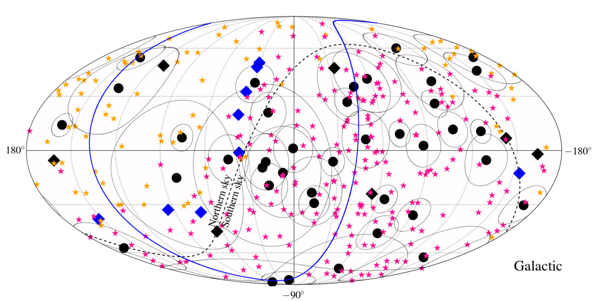
Point-source searches

in Astro-Particle Physics

Charles Timmermans

Nikhef, Radboud University



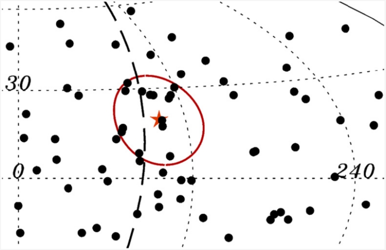
Literature

- Ti-Pei Li and Yu-Qian Ma, Astrophys.J. 272 (1983)
- Telescope Array Collaboration
  - Astrophys.J. 804 (2015)
  - Astrophys.J. 790 (2014)
- J. Brown *et al.*, Astropart.Phys. 33 (2010)
- IceCube Collaboration
  - Astrophys.J. 796 (2014)
  - Astrophys.J. 779 (2013)
- Auger Collaboration
  - Astrophys.J. 804 (2015)
  - Phys.Rev.Lett. 101 (2008)
- TA, IC, Auger collaborations, JCAP 1601 (2016)

- Specific problems for searches in APP
  - source location unknown
  - source characteristics unknown
  - background model unreliable
  - data volume grows linearly
- Single source and Prescription (Auger correlation)
- Significances (TA enhancement)
- Unbinned likelihood (Icecube search)

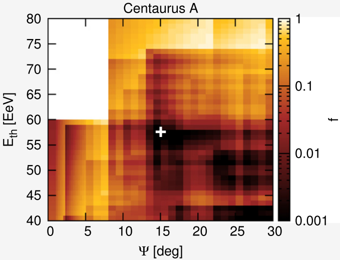
Centaurus A: The Auger warm spot

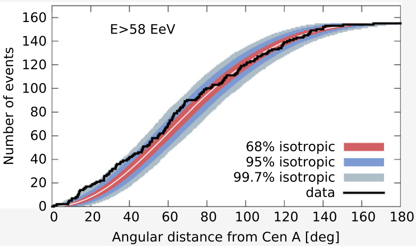
- Event selection:
  - $E > 50\,\text{EeV}$
  - $\psi < 18^\circ$



- Situation in 2010:
  - 13 events correlate within 18° where 3.2 are expected
  - Poisson probability $P(N \geq 18, 3.2) \simeq 10^{-4}$
- Situation in 2015:
  - 18 events correlate where 9 are expected

Centaurus A: The Auger warm spot





- 14 events with expectation of 4.2: $p = 2 \cdot 10^{-4}$
  - scan over many not-independent possibilities
  - How to calculate the significance?
  - simulate same number of events as data isotropically distributed many times
- For Centaurus A $\mathcal{P} \simeq 1.4\,\%$

Catalog correlation

In 2007 Auger investigated the correlation of UHECR with objects from the VCV catalog.
Chance probability of correlation of at least k uniformly distributed cosmic rays:

$$P = \sum_{j=k}^{N} \frac{N!}{(N-j)!j!} p^j (1-p)^{N-j}$$

p is the chance probability of cosmic ray to be correlated with catalog.
p determined from MC.

Catalog correlation

Problem: Limited dataset that does not grow exponentially.
Solution: Scan parameters with available data, use new data to test this *prescription*

Dataset: all obtained before May 27, 2006.
Selection

- Source candidates within 75 Mpc
- Event energy > 56 EeV
- Angular distance to source within 3.1°
- p = 21 %
- 12 of 15 events correlate to catalog.
- 3.2 events expected.

Test Criterium

Null hypothesis: Data is uniformly distributed

- α: probability to reject null hypothesis incorrectly
  - Declaring anisotropy when it is not
  - $\alpha = 0.01$
- β: probability to accept null hypothesis incorrectly
  - Declare isotropy when it is not
  - $\beta = 0.05$
- Perform a running prescription
  - Reject isotropy with at least $(1 - \alpha) = 0.99$
  - Minimal correlation 60 %
  - β remains below 5 %

Test Criterium

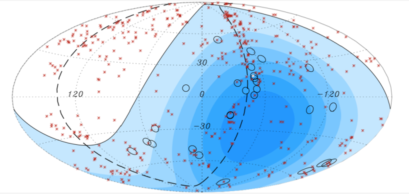
A positive exit for new data (Binomial Statistics):

N	4	6	8	10	12	...	30	31	33	34
k_{min}	4	5	6	7	8	...	14	14	15	15

- May 25 2007 **6 out of 8** events correlated.
  - Prescription satisfied
- August 31 2007 **8 out of 13** events correlated
  - Chance probability $1.7 \cdot 10^{-3}$

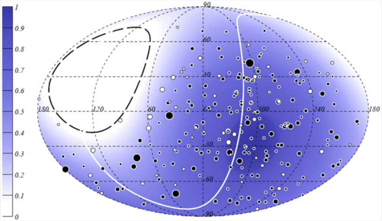
Result

2007



First series of events correlating.

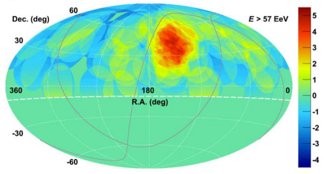
2015



Correlation disappeared.

Note that the probability to reject the null hypothesis incorrectly was 1 %!

Telescope Array excess



Significance of local excess: 5.1
What does this mean?

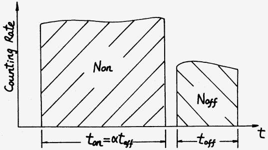
Luckily the paper answers the question:
We calculated the statistical significance of the excess of events compared to the background events at each grid point of the sky using the following equation (Li & Ma 1983)

$$S_{LM} = \sqrt{2}\left[N_{on}\ln\frac{(1+\alpha)N_{on}}{\alpha(N_{on}+N_{off})} + N_{off}\ln\frac{(1+\alpha)N_{off}}{N_{on}+N_{off}}\right]^{\frac{1}{2}}$$

Lets discuss Li-Ma significance first.

Li-Ma Significance

First used in gamma-astronomy



$$\alpha = \frac{T_{on}}{T_{off}}$$
$$\hat{N}_B = \alpha N_{off}$$
$$N_S = N_{on} - \hat{N}_B = N_{on} - \alpha N_{off}$$

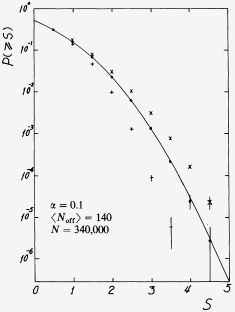
What is the significance of an enhancement of the measurement in a signal region?

Assumptions:

- Background determined from a measurement
    - Thus contains uncertainty
- Signal region contains a background fluctuation

Li-Ma Significance

Frequentist interpretation



$$S = \frac{N_S}{\hat{\sigma}(N_s)} = \frac{N_{on} - \alpha N_{off}}{\sqrt{N_{on} + \alpha^2 N_{off}}}$$

Line: Normal distribution
Plusses: Significance S

Problem: The calculated significance is not in agreement with a background only hypothesis.

Solution: Make this hypothesis explicit!

Li-Ma Significance

Assuming Background only (Null hypothesis):
also in the signal bin, there is only background. The best background estimate includes the signal bin:

$$\hat{N}_B = \frac{N_{on} + N_{off}}{t_{on} + t_{off}} t_{on} = \frac{\alpha}{1+\alpha}\left(N_{on} + N_{off}\right)$$

Li-Ma Significance

Suppose we have data $X(x_1, ..x_N)$, and parameters $E(\epsilon_1 ... \epsilon_M)$ and statistical hypotheses E_0 and $E \neq E_0$, we can define the maximum likelihood ratio

$$\lambda = \frac{L(X|E_0)}{L(X|\hat{E})} = \frac{P_r(X|E_0)}{P_r(X|\hat{E})}$$

If the null hypothesis is true, $-2\ln\lambda$ follows a χ^2 distribution with r degrees of freedom.
Here $X = (N_{on}, N_{off})$, $E = (N_S, N_B)$, $E_0 = (0, N_B)$

In this case only N_S is involved in the null hypothesis, so $r = 1$.

Li-Ma Significance

The maximum likelihood:

$$\begin{aligned} L(X|E) &= P_r(N_{on}, N_{off}|N_S, N_B) \\ &= P_r(N_{on}, N_{off}|N_{on} - \alpha N_{off}, \alpha N_{off}) \\ &= P_{Poisson}(N_{on}|N_{on}) P_{Poisson}(N_{off}|N_{off}) \\ &= \frac{N_{on}^{N_{on}}}{N_{on}!} e^{-N_{on}} \frac{N_{off}^{N_{off}}}{N_{off}!} e^{-N_{off}} \end{aligned}$$

Keep in mind:

$$P_{Poisson}(N|\mu) = \frac{\mu^N}{N!} e^{-\mu}$$

Li-Ma Significance

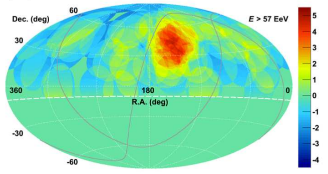
$$
\begin{aligned}
L(X|E_0) &= P_r(N_{on}, N_{off}|N_S, N_B) \\
&= P_r(N_{on}, N_{off}|0, \frac{\alpha(N_{on}+N_{off})}{1+\alpha}) \\
&= P_{Poisson}(N_{on}|\frac{\alpha(N_{on}+N_{off})}{1+\alpha}) P_{Poisson}(N_{off}|\frac{(N_{on}+N_{off})}{1+\alpha})
\end{aligned}
$$

Therefore:

$$\lambda = \left[\frac{\alpha}{1+\alpha}\frac{N_{on}+N_{off}}{N_{on}}\right]^{N_{on}} \left[\frac{1}{1+\alpha}\frac{N_{on}+N_{off}}{N_{off}}\right]^{N_{off}}$$

$$S_{LM} = \sqrt{2}\left[N_{on}\ln\frac{(1+\alpha)N_{on}}{\alpha(N_{on}+N_{off}} + N_{off}\ln\frac{(1+\alpha)N_{off}}{N_{on}+N_{off}}\right]^{\frac{1}{2}}$$

Telescope Array excess

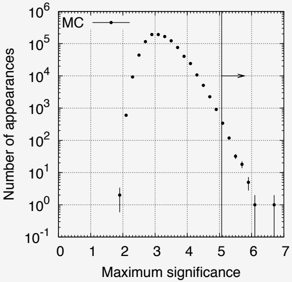


A local excess of 5.1 sigma is observed.
The probability of this to happen **in any test** is $3.4 \cdot 10^{-7}$

How many tests have been performed? The sky is divided into many bins, all correlated!

Telescope Array excess

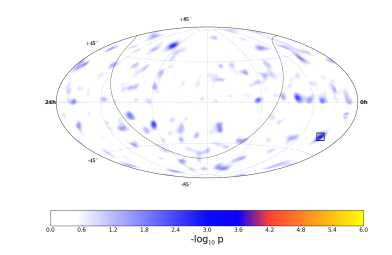
- Generate MC. 1,000,000 MC datasets are generated.
- Perform same analysis
- Safe Maximum Significance



In 365 out of 1,000,000 datasets
$S_{LM} > 5.1$
Thus $P = 3.7 \cdot 10^{-4}$.

IceCube Discovery Potential

Searches for Extended and Point like sources using IceCube



The significance is estimated by using an unbinned likelihood ratio test (Braun *et al.* 2010).

Unbinned likelihood

- $\mathcal{S}_i$: Signal PDF
- $\mathcal{B}_i$: Background PDF
- n_s: Unknown contribution of signal events

$$\mathcal{L}_i(n_s) = \frac{n_s}{N}\mathcal{S}_i + \left(1 - \frac{n_s}{N}\right)\mathcal{B}_i$$

And for all events:

$$\mathcal{L}(n_s) = \prod_{i=1}^{N}\left[\frac{n_s}{N}\mathcal{S}_i + \left(1 - \frac{n_s}{N}\right)\mathcal{B}_i\right]$$

Likelihood ratio test:

$$D = -2\ln(\lambda) = -2\ln\left[\frac{\mathcal{L}(n_s = 0)}{\mathcal{L}(\hat{n}_s)}\right] \times \text{sign}(\hat{n}_s)$$

A time integrated search

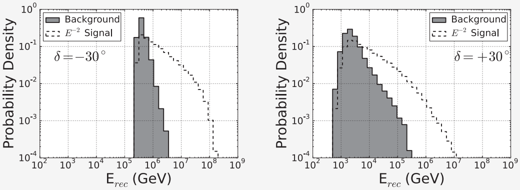
$$\mathcal{S}_i = S_i(|\vec{x}_i - \vec{x}_s|, \sigma_i)\mathcal{E}_i(E_i, \delta_i, \gamma)$$

where

$$S_i = \frac{1}{2\pi\sigma_i} e^{-\frac{|\vec{x}_i - \vec{x}_s|^2}{2\sigma_i^2}}$$

The energy dependence takes into account that the signal source is assumed to be harder than the background:

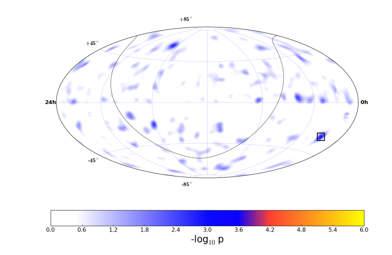
- Background: $\frac{dN}{dE} \sim E^{-3.6}$
- Source: $\frac{dN}{dE} \sim E^{-\gamma}$, where γ = 2.0, 2.3, 2.6 ...
- Energy response depends upon declination



A time integrated search

$$\mathcal{B}_i = \frac{1}{\Omega(\delta_i)} \mathcal{E}_i(E_i, \delta_i, \mathrm{ATM}_\nu)$$

where $\Omega(\delta_i)$ is the solid angle centered around the declination band.



The local significance is then given as a probability (frequentist approach) of obtaining at least the likelihood ratio found.

Using an ensemble of scrambled data the trial factor (look elsewhere effect) is taken into account. All significances $> 10\,\%$

Correlation IceCube and Auger/TA

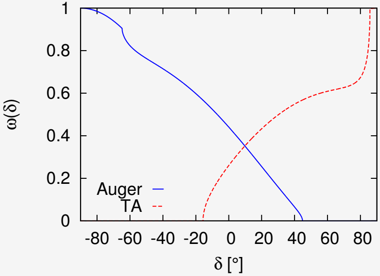
Assume IceCube neutrino's point to sources of Cosmic Rays.

$$\log \mathcal{L}(n_s) = \sum_{i=1}^{N_{\text{Auger}}} \left[\frac{n_s}{N_{CR}} \mathcal{S}_i^{\text{Auger}} + \left(1 - \frac{n_s}{N_{CR}}\right) \mathcal{B}_i^{\text{Auger}} \right] + \sum_{i=1}^{N_{\text{TA}}} \left[\frac{n_s}{N_{CR}} \mathcal{S}_i^{\text{TA}} + \left(1 - \frac{n_s}{N_{CR}}\right) \mathcal{B}_i^{\text{TA}} \right]$$

Signal PDF:

$$\mathcal{S}_i^{\text{Auger}}(\vec{r}_i, E_i) = R_{\text{Auger}}(\delta_i) \sum_{j=1}^{N_{\text{src}}} S_j(\vec{r}_i, \sigma(E_i))$$

Correlation IceCube and Auger/TA



$$S_j = \frac{1}{2\pi\sigma_i} e^{-\frac{|\vec{r}_i - \vec{s}_S|^2}{2\sigma_i^2}}$$

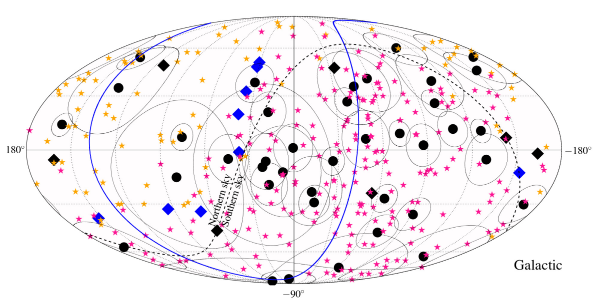
$$\sigma_i^2 = \sigma^2_{\mathrm{Auger_TA}} + \sigma^2_{\mathrm{MD}}(E_i)$$

$$\sigma_{\mathrm{MD}}(E) = D \times 100/E$$

$R_{\mathrm{Auger,TA}}(\delta_i)$

Background likelihood: follows exposure of Auger or TA.
Perform likelihood ratio test as before.

Result Correlation IceCube and Auger/TA



	High-energy tracks			High-energy cascades		
D	n_s	TS	pre-trial p-value	n_s	TS	pre-trial p-value
3°	4.2	0.6	0.22	53.7	8.21	2.1×10^{-3}
6°	0.5	2.7×10^{-3}	0.48	85.7	11.99	2.7×10^{-4}
9°	0	0	under-fluctuation	106.1	11.32	3.8×10^{-4}

Conclusion

- Li-Ma significance used to calculate local probability
  - when background estimated from data
- Many datasets generated in order to assess global probability
  - Frequentist approach to probability
- Data volume grows linearly, truly blind data is still to come
  - prescription approach cannot be too strict