Multivariate analysis approach to ultra high energy cosmic ray mass composition

Gašper Kukec Mezek[1], prof. dr. Andrej Filipčič[1,2]

[1]University of Nova Gorica, [2]Jozef Stefan Institute

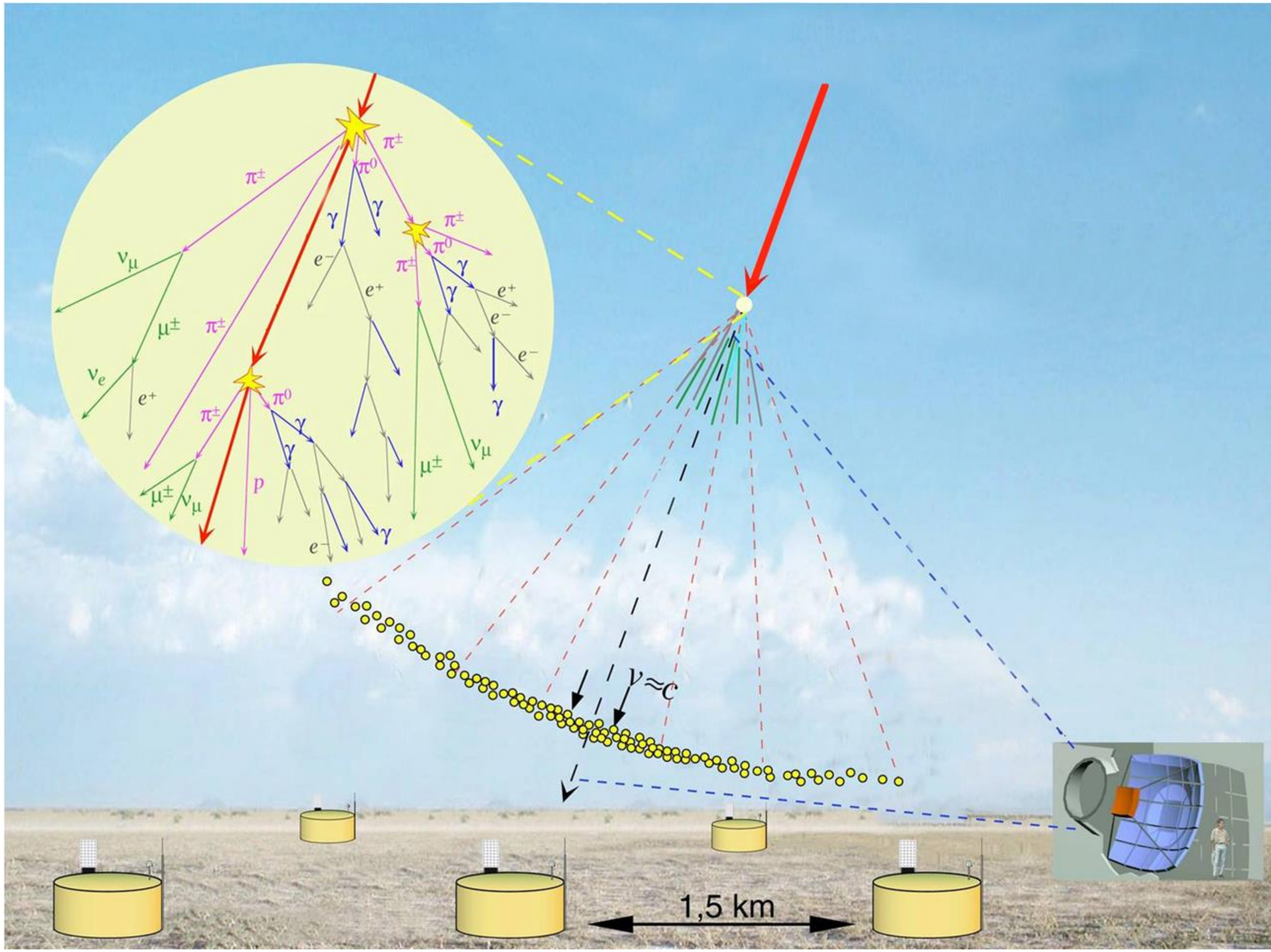
AstroTS (September 25th 2017)

Outline

- Ultra high energy cosmic rays (UHECRs)
- Detection with Pierre Auger Observatory
- Why cosmic ray mass composition?
- Mass sensitive observables
  - Depth of shower maximum (X_{max})
  - Surface detector signal at 1000m from the shower axis (S_{1000})
  - Risetime of surface detector signal at 1000m from the shower axis ($t_{1/2}$)
- Mass composition results
- Multivariate analysis (MVA)
  - Multivariate analysis of observables
  - Preliminary analysis
- Summary and outlook

Ultra high energy cosmic rays (UHECRs)

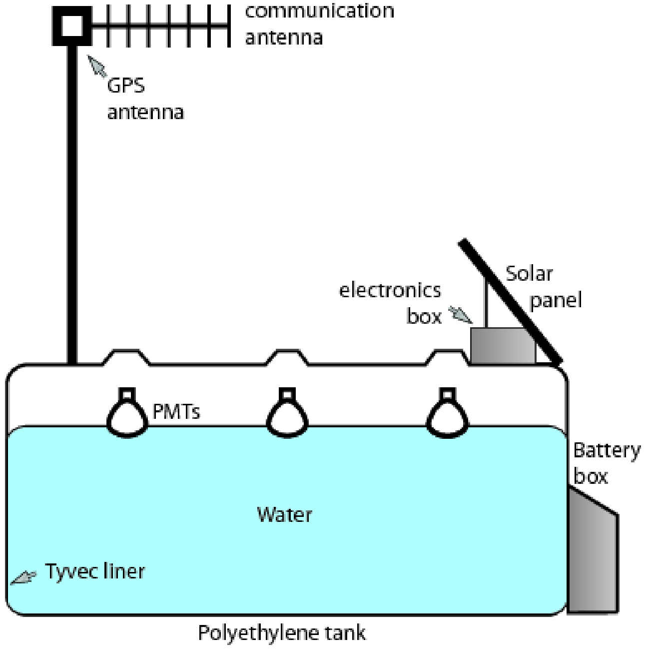
- Cosmic rays (CRs): Particles that arrive to Earth from space
- Ultra high energy cosmic rays (UHECRs): CRs with energies above $\sim 10^{18} eV$
- Extensive air showers (EAS): Cascade of particles in the atmosphere caused by interaction of CRs and atmospheric nuclei



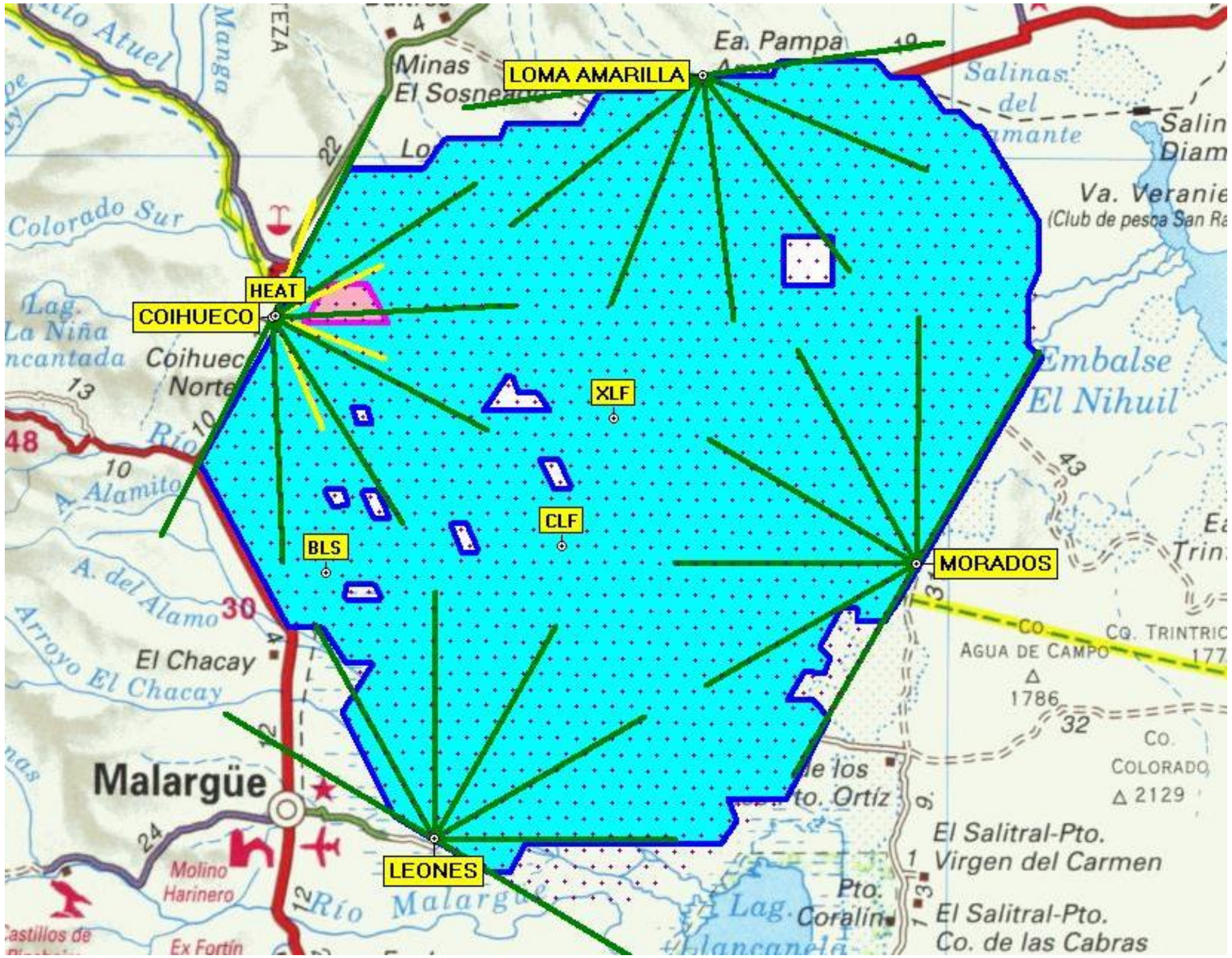
[apcauger.in2p3.fr/Public/Presentation/index.php]

Detection with Pierre Auger Observatory

- Hybrid detection system:
  - Water Cherenkov stations (SD)
  - Fluorescence telescopes (FD)
- 1600 SD stations covering $\sim 3000\ km^2$
- Filled with water to observe Cherenkov light with 3 PMTs



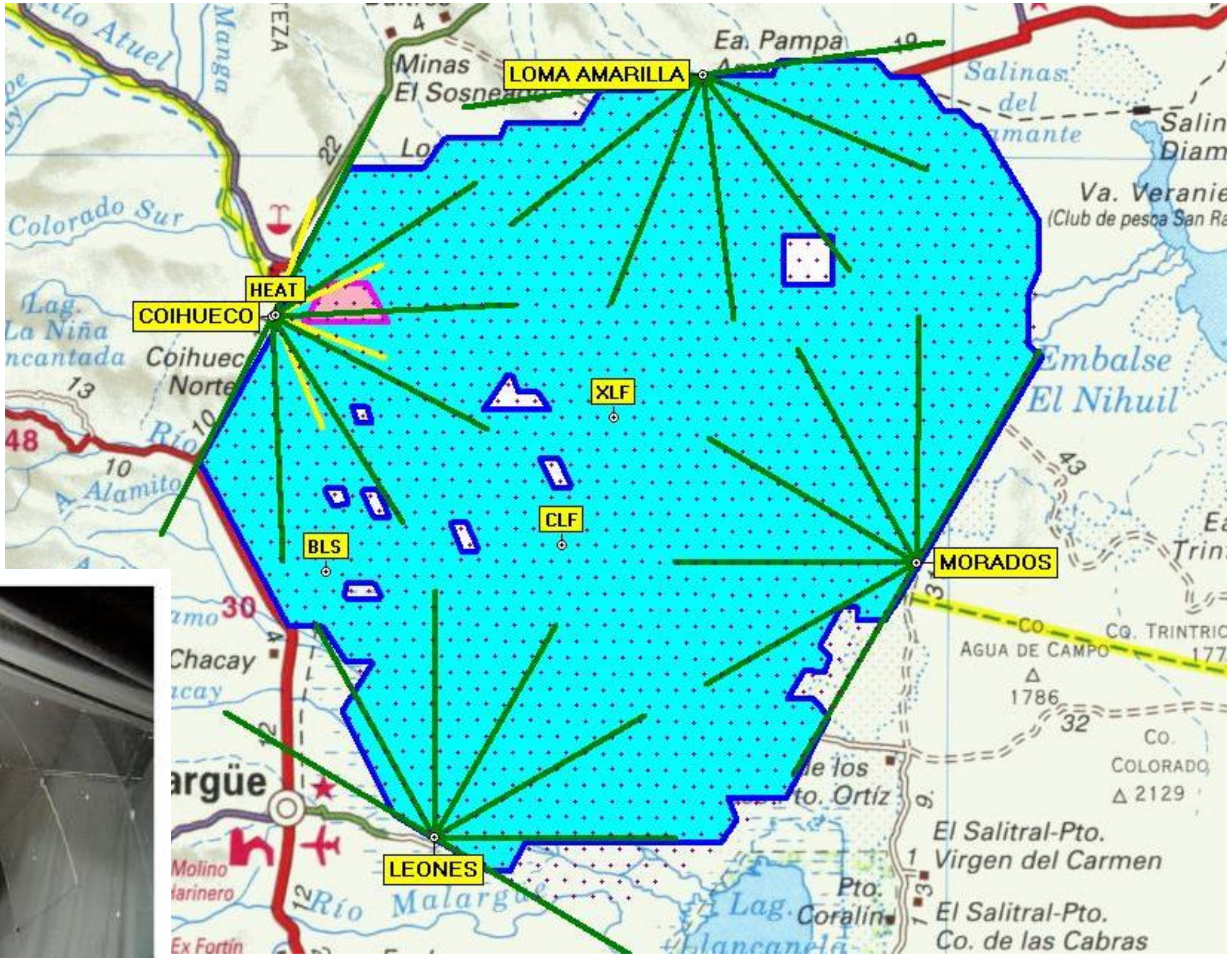
[inspirehep.net/record/1123321/files/Tanks.png]



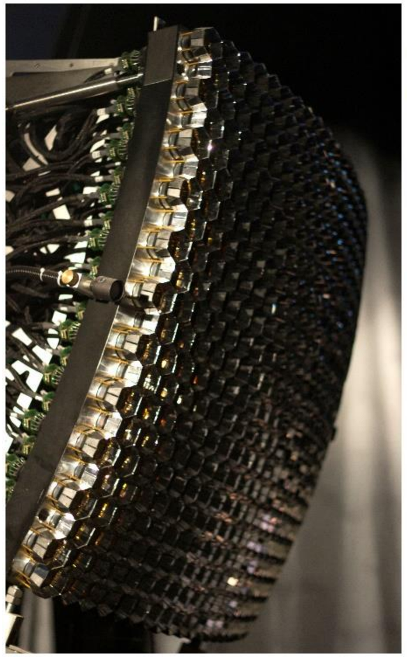
[www.auger.org]

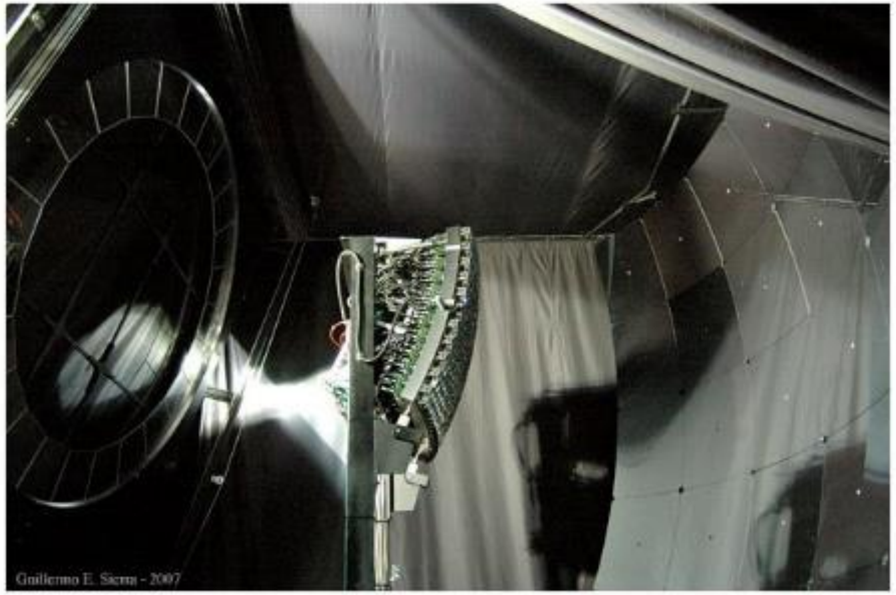
Detection with Pierre Auger Observatory

- Hybrid detection system:
  - Water Cherenkov stations (SD)
  - Fluorescence telescopes (FD)
- 4 FD detectors with 6 telescopes each (+HEAT with 3 telescopes)
- Observing UV light from N_2 excitations



[www.auger.org]



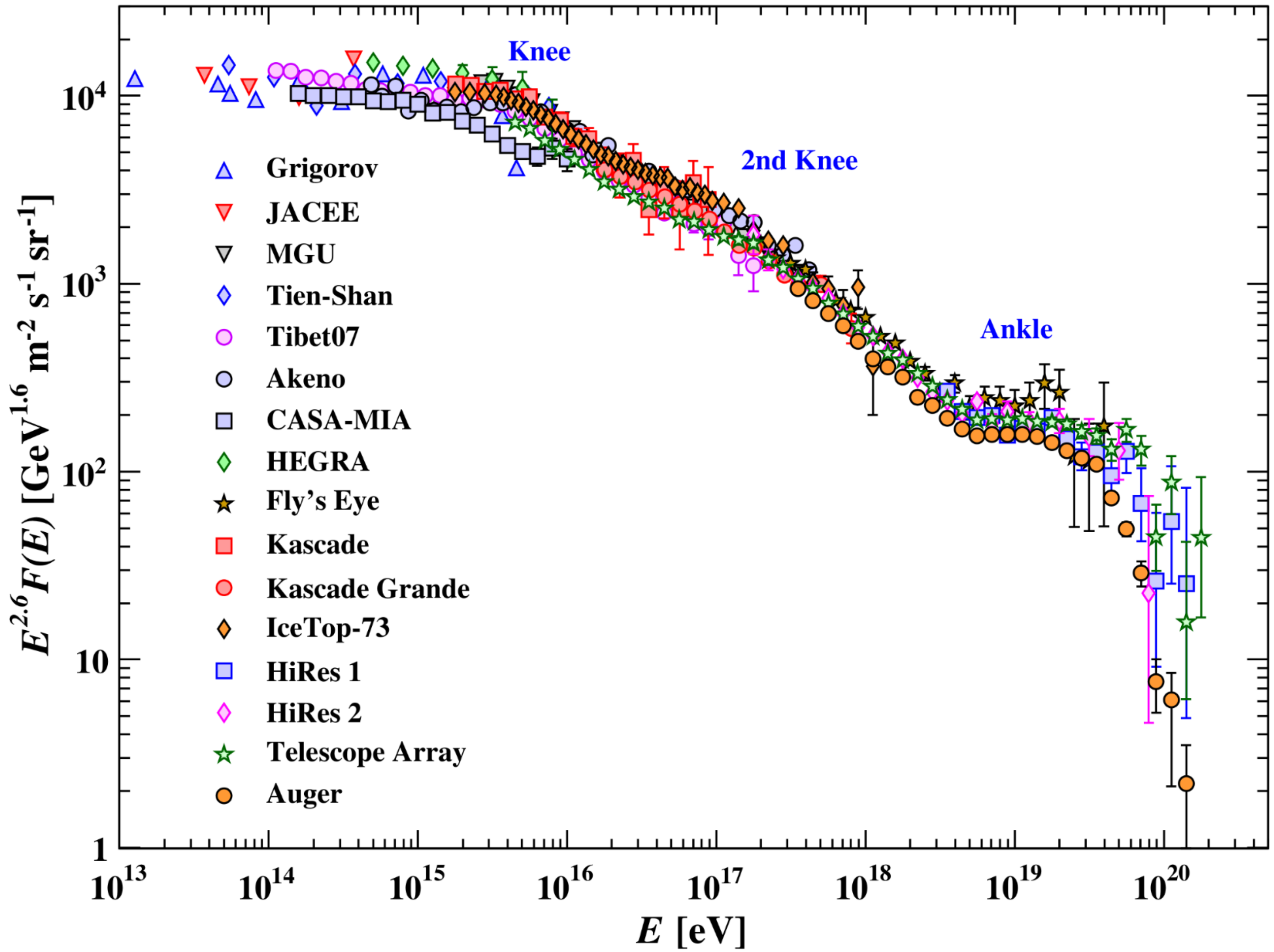


[G. Sierra, www.auger.org]

Why cosmic ray mass composition?

- High-energy cosmic rays (CRs) are messengers of violent astrophysical phenomena
- CR mass composition still unclear at highest energies
- Why identify high energy cosmic rays?
  - Access to energies higher than at any man-made colliders (LHC)
  - Backtracking to their acceleration origin (extragalactic)
  - Information on acceleration processes, magnetic field strength,…

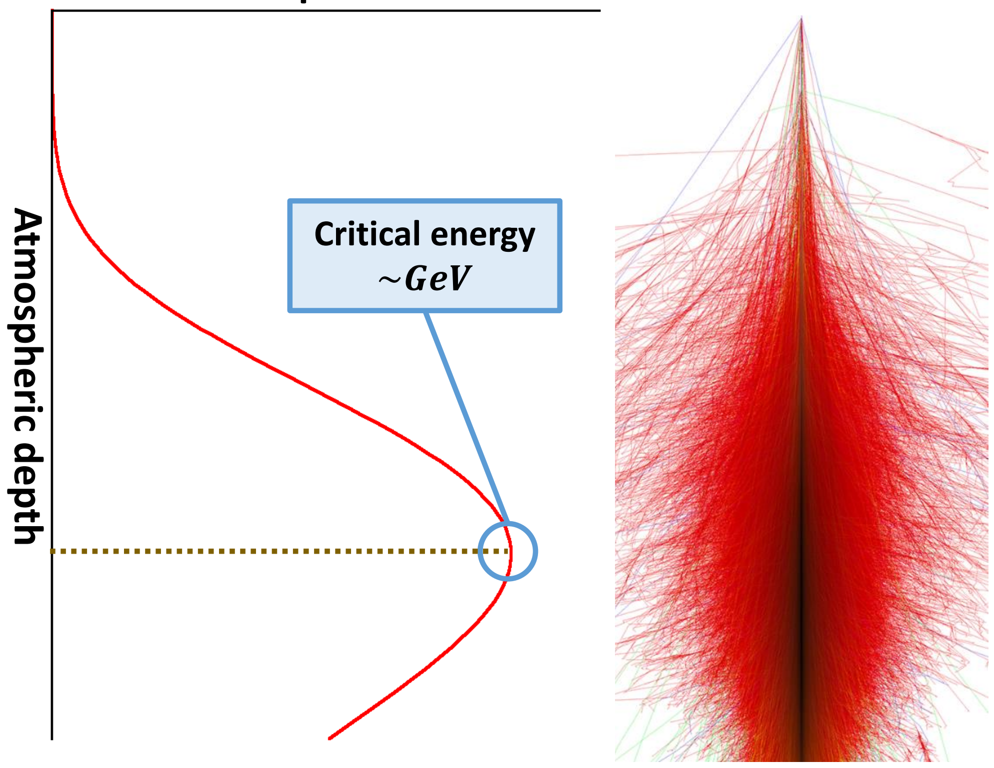
[C. Patrignani et al. (Particle Data Group), Chin. Phys. C 40, 100001 (2016)]



Mass sensitive observables

Mass sensitive observables

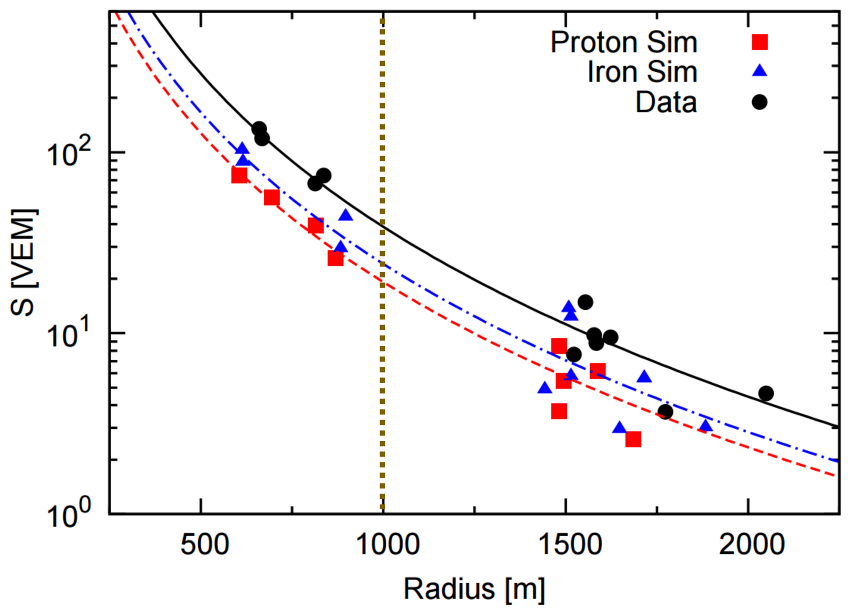
[www-zeuthen.desy.de/~jknapp/fs/photon-showers.html]



Depth at shower maximum (X_{max})

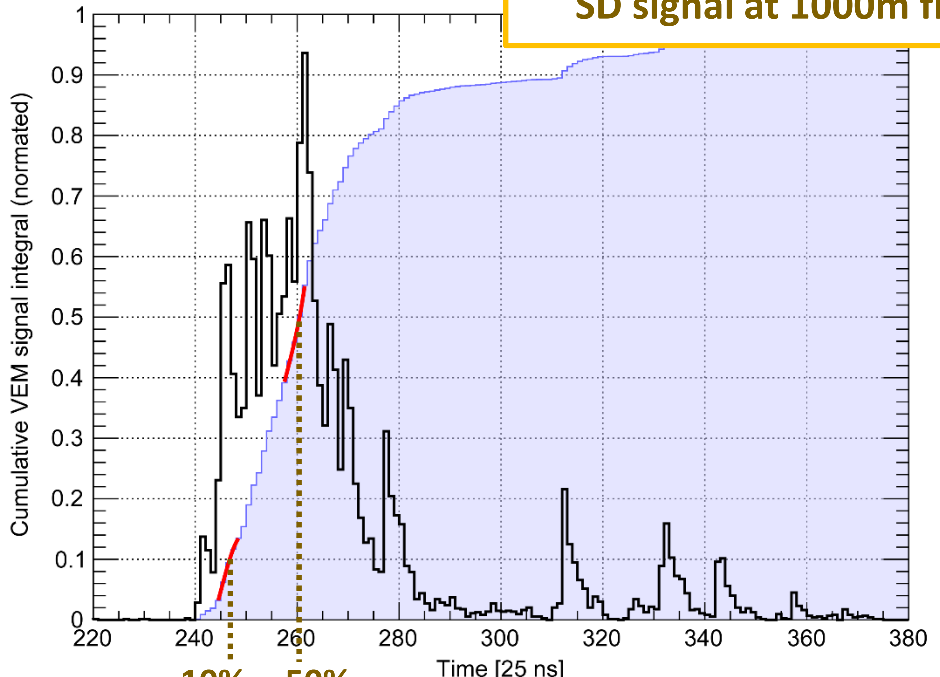
Heavy < Light

[J. Allen et al., ICRC 2011, arXiv:1107.4804]



SD signal at 1000m from shower axis (S_{1000})

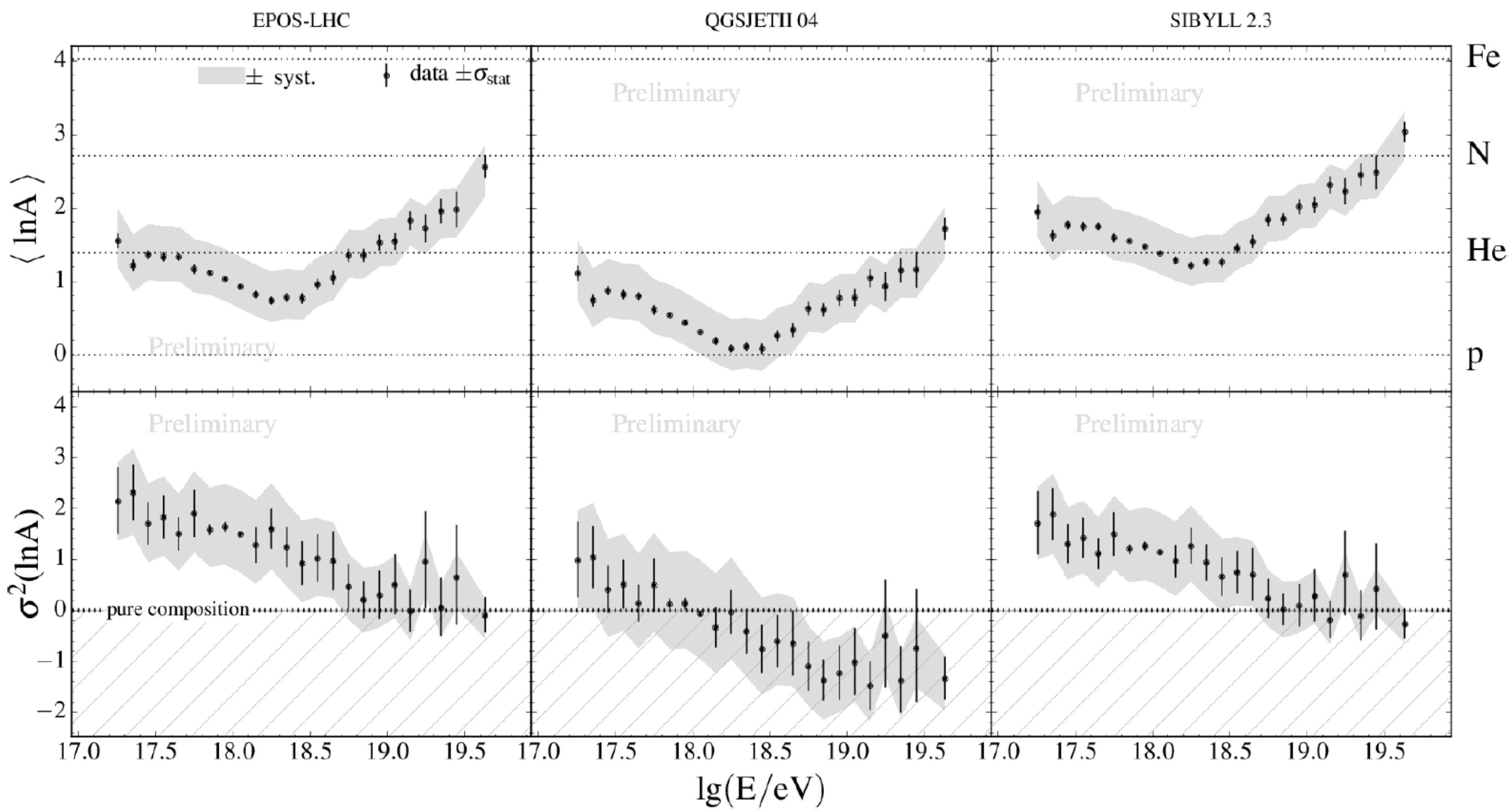
Heavy > Light

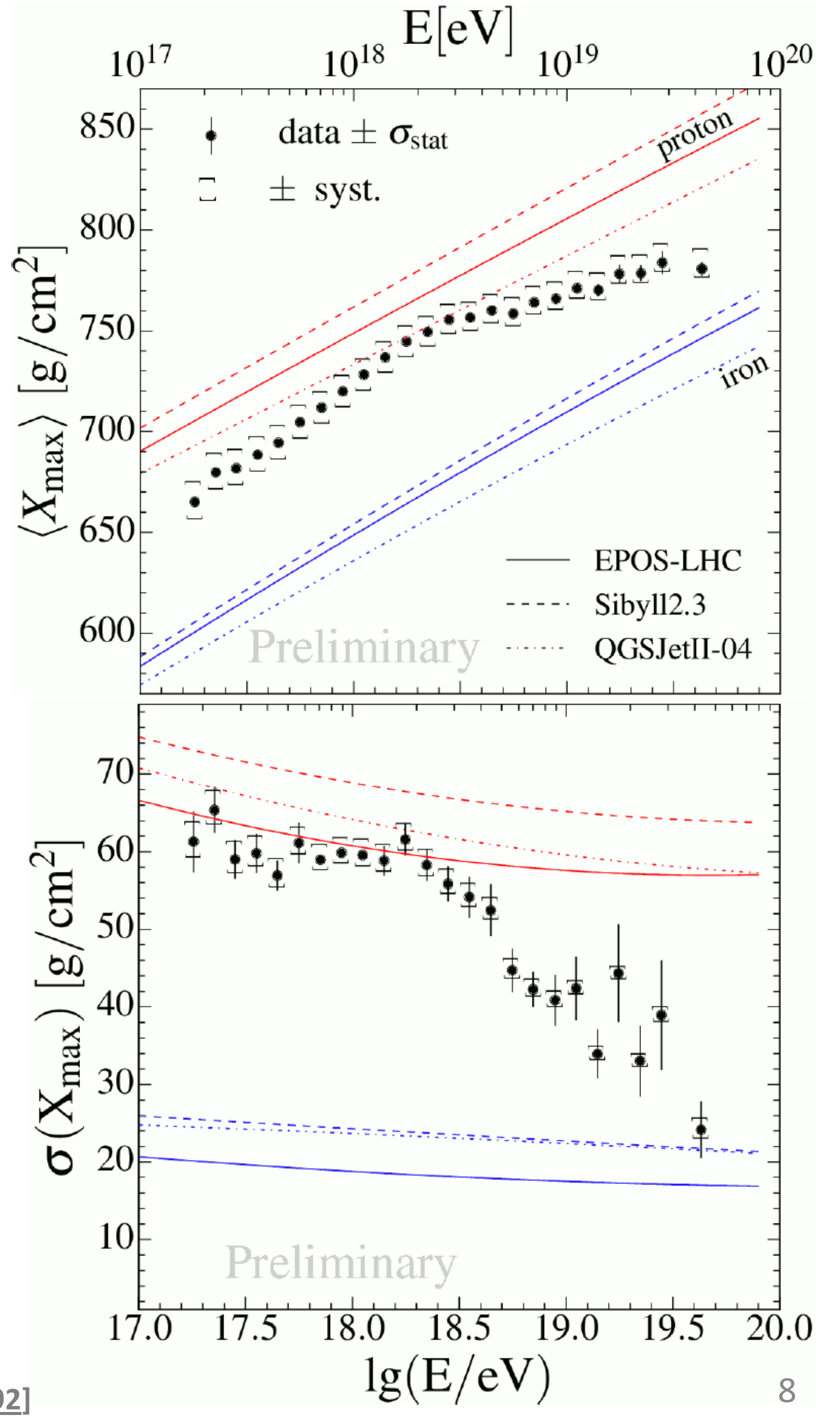


Risetime at 1000m from shower axis ($t_{1/2}$)

Heavy < Light

Mass composition results

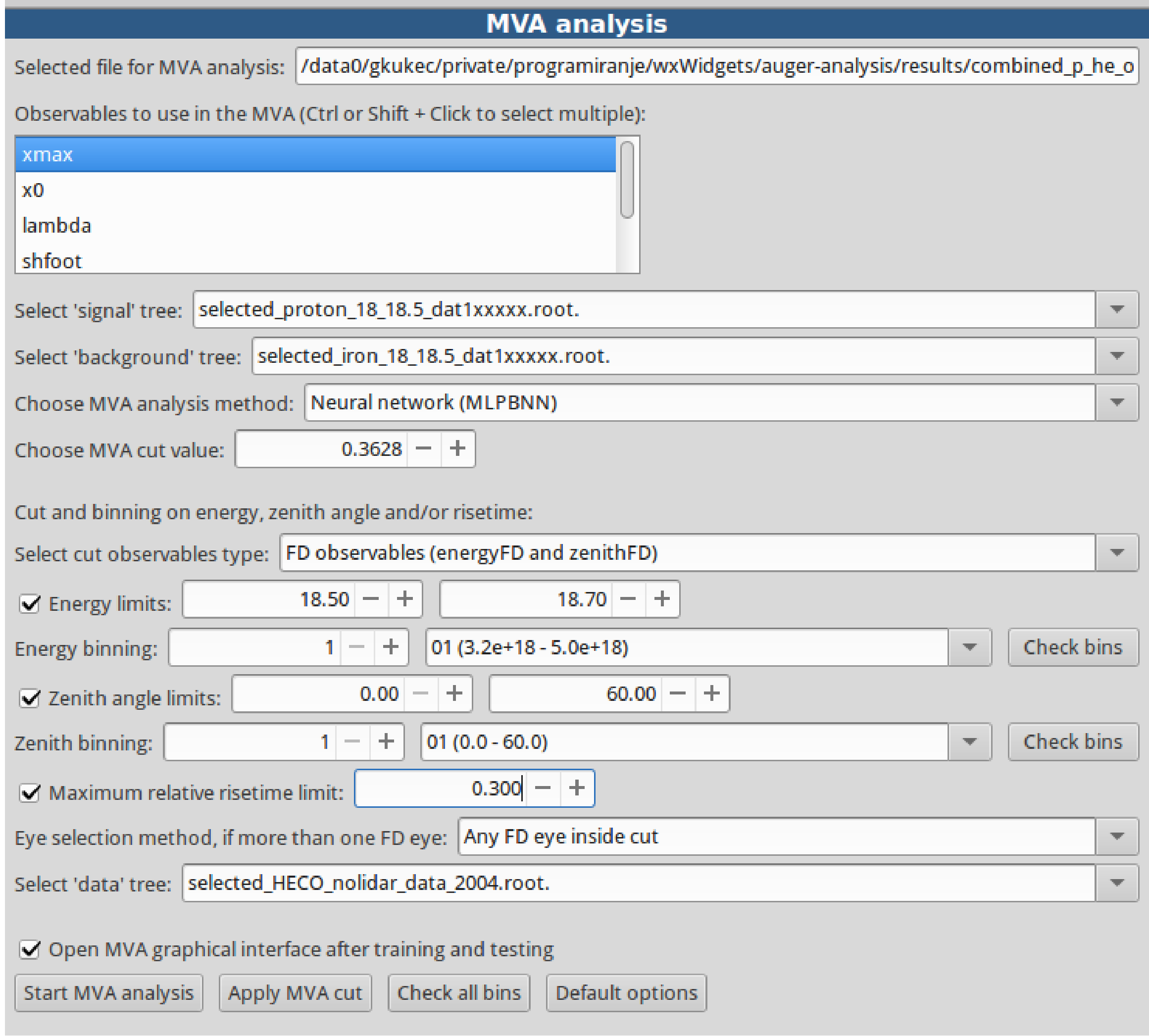




[J. Bellido for the Pierre Auger Collaboration, ICRC 2017, arXiv:1708.06592]

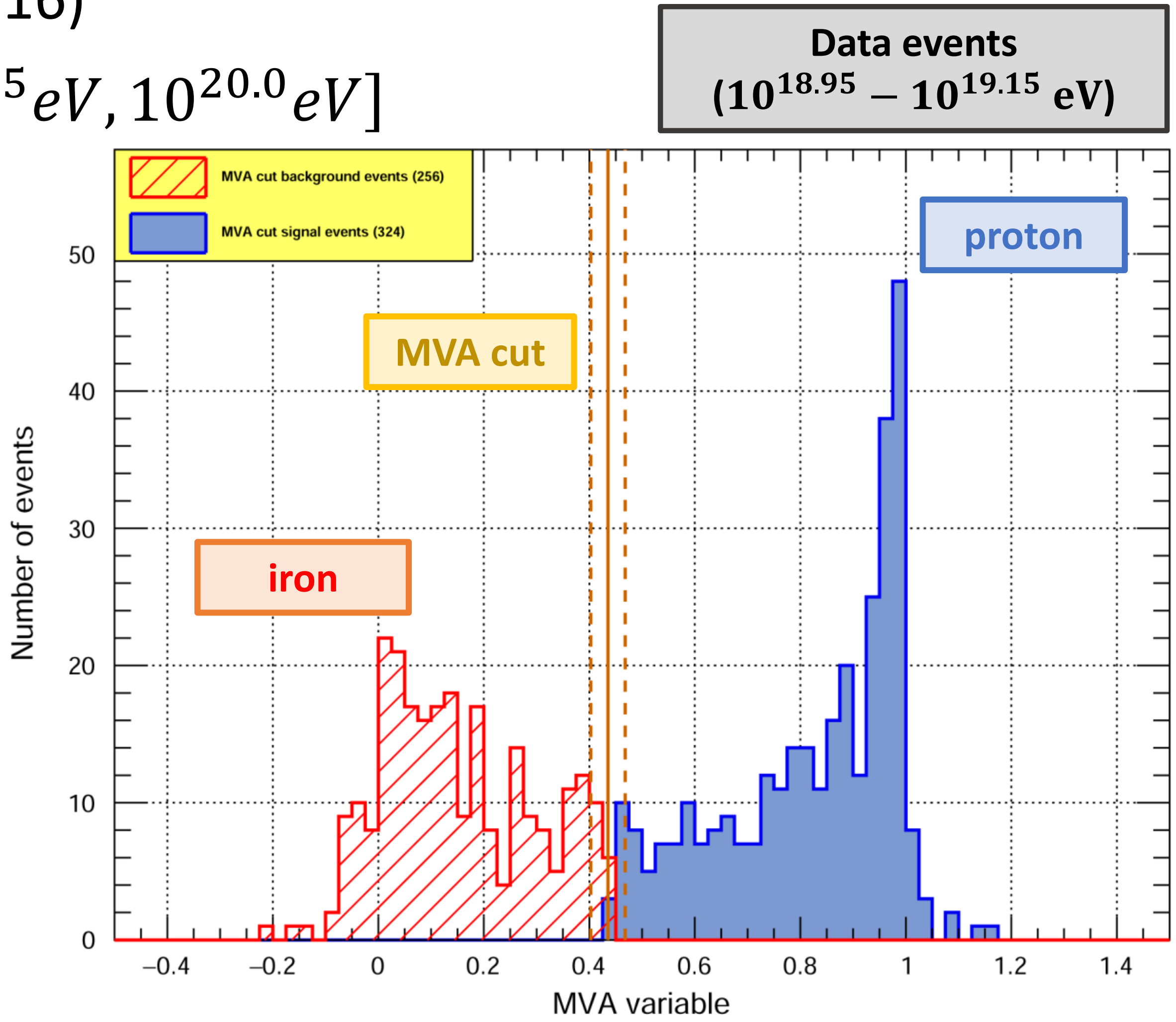
Multivariate analysis (MVA)

- Statistical technique to analyse data, dependent on more than one variable
- Analysis includes machine learning algorithms with artificial neural networks – not widely used in astroparticle physics
- Mass composition studies mostly done on a small number of observables or individually (ex. Phys. Rev. D 90, 122006 (2014), arXiv:1708.06592)
- Using MVA of more observables could give better mass composition estimation
- Custom analysis software interfacing TMVA package from ROOT and Pierre Auger Observatory file structure

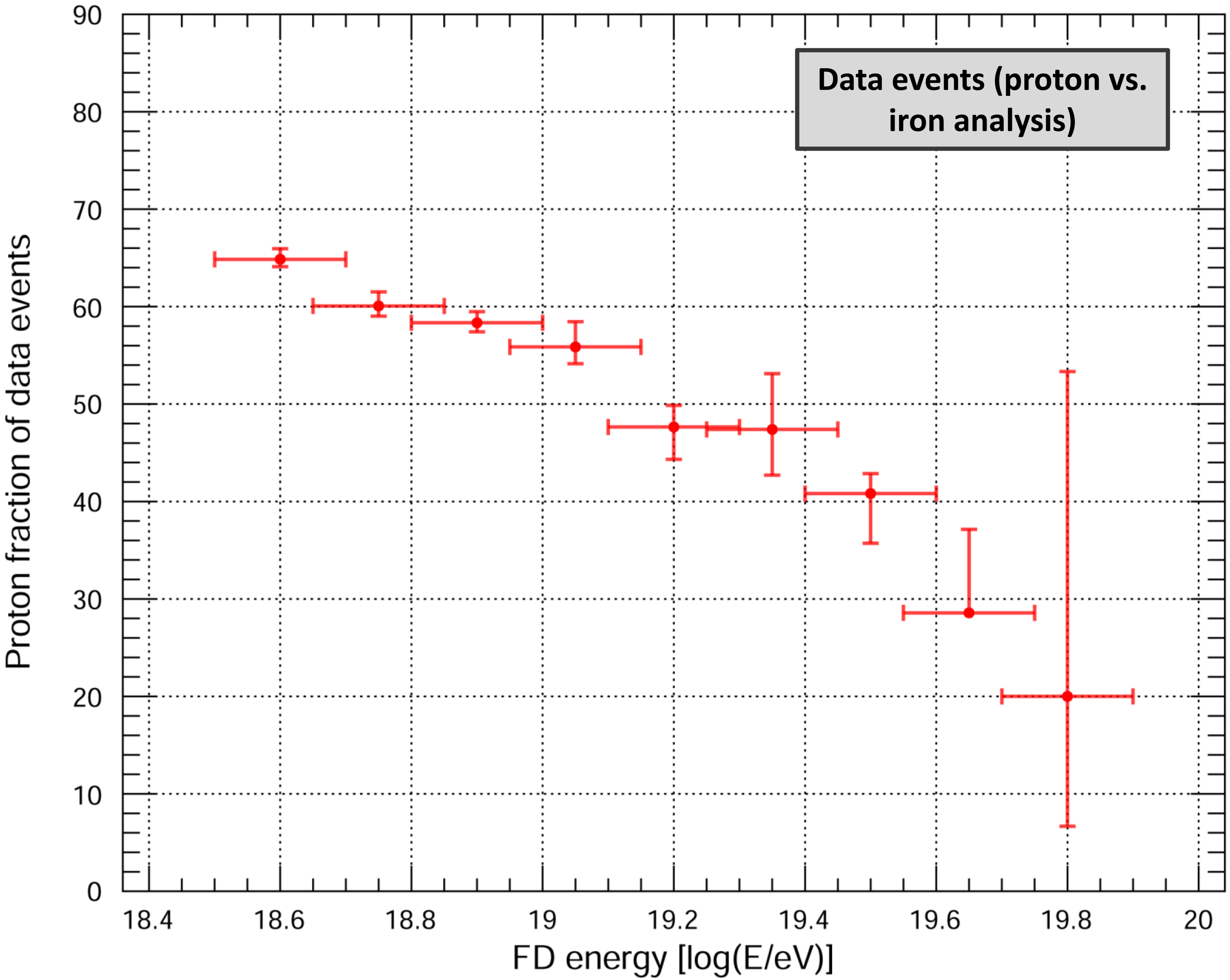


Multivariate analysis (MVA)

- Combine multiple mass sensitive observables into a single MVA variable
- Simulations from Napoli shower library (http://natter.na.infn.it:18501), data from Pierre Auger Observatory (2004 – 2016)
- Selecting hybrid events with $E_{FD} = [10^{18.5} eV, 10^{20.0} eV]$ and $\theta_{FD} = [0°, 60°]$
- Stages of MVA:
  1. Divide events into energy subsets
  2. Choose two different primary particle simulation sets as „signal“ and „background“ events
  3. Test and train selected MVA method to get the MVA variable (ROOT TMVA)
  4. Create a cut on the MVA variable and apply it to simulation (purity) and data events (application)



Preliminary analysis



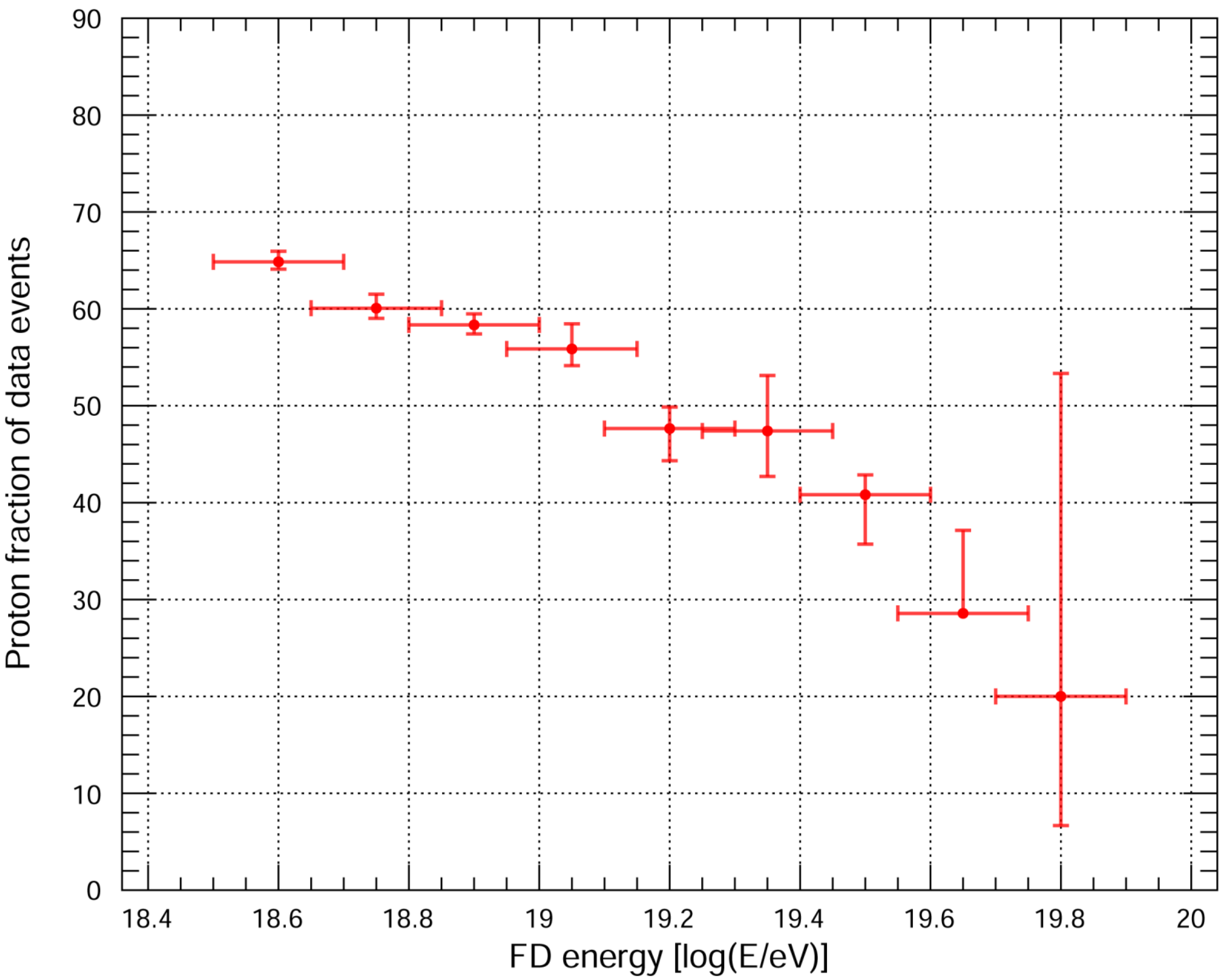
Energy bin ($\log E/eV$)	Number of data events
18.5 – 18.7	2649
18.65 – 18.85	1525
18.8 – 19.0	965
18.95 – 19.15	580
19.1 – 19.3	361
19.25 – 19.45	192
19.4 – 19.6	98
19.55 – 19.75	35
19.7 – 19.9	15

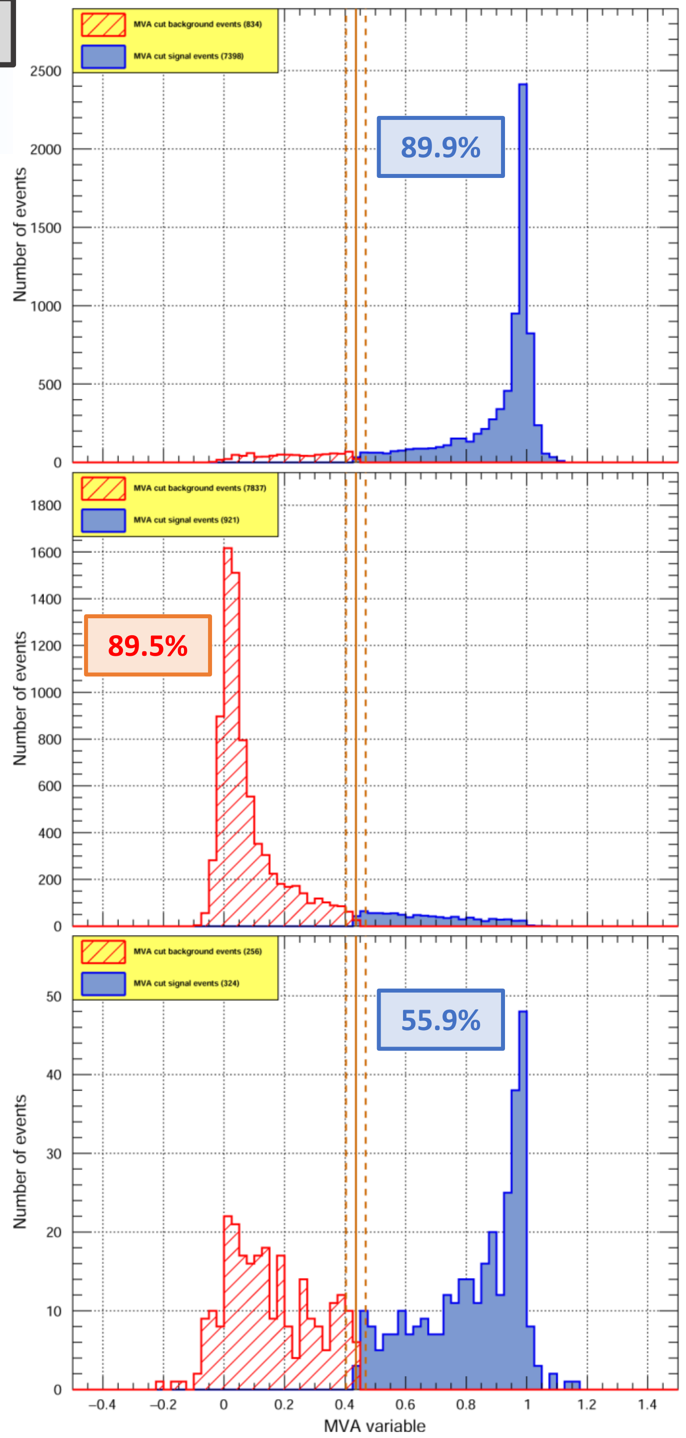
Preliminary analysis

Multivariate analysis (MVA)

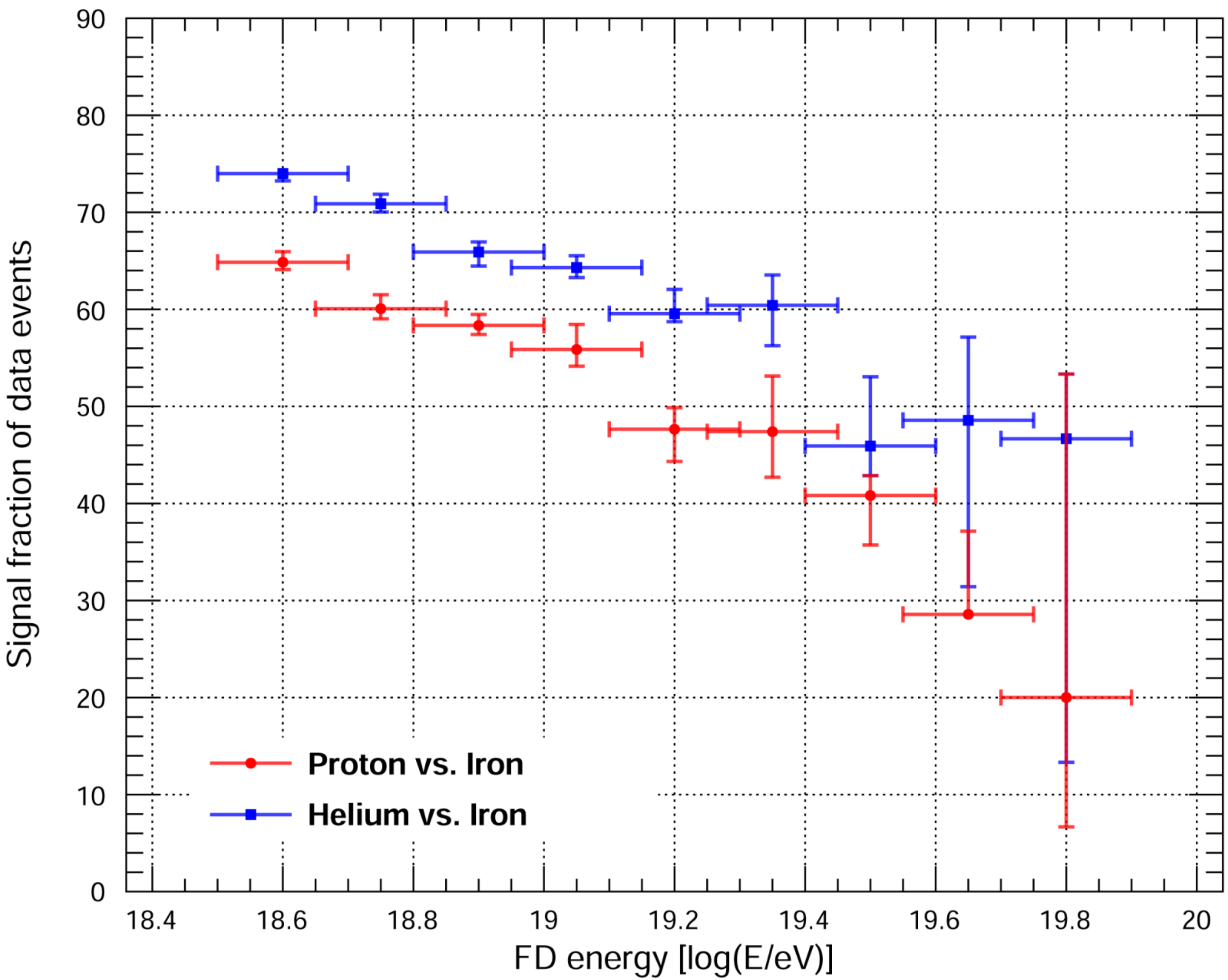
Preliminary analysis

$10^{18.95} - 10^{19.15}$ eV

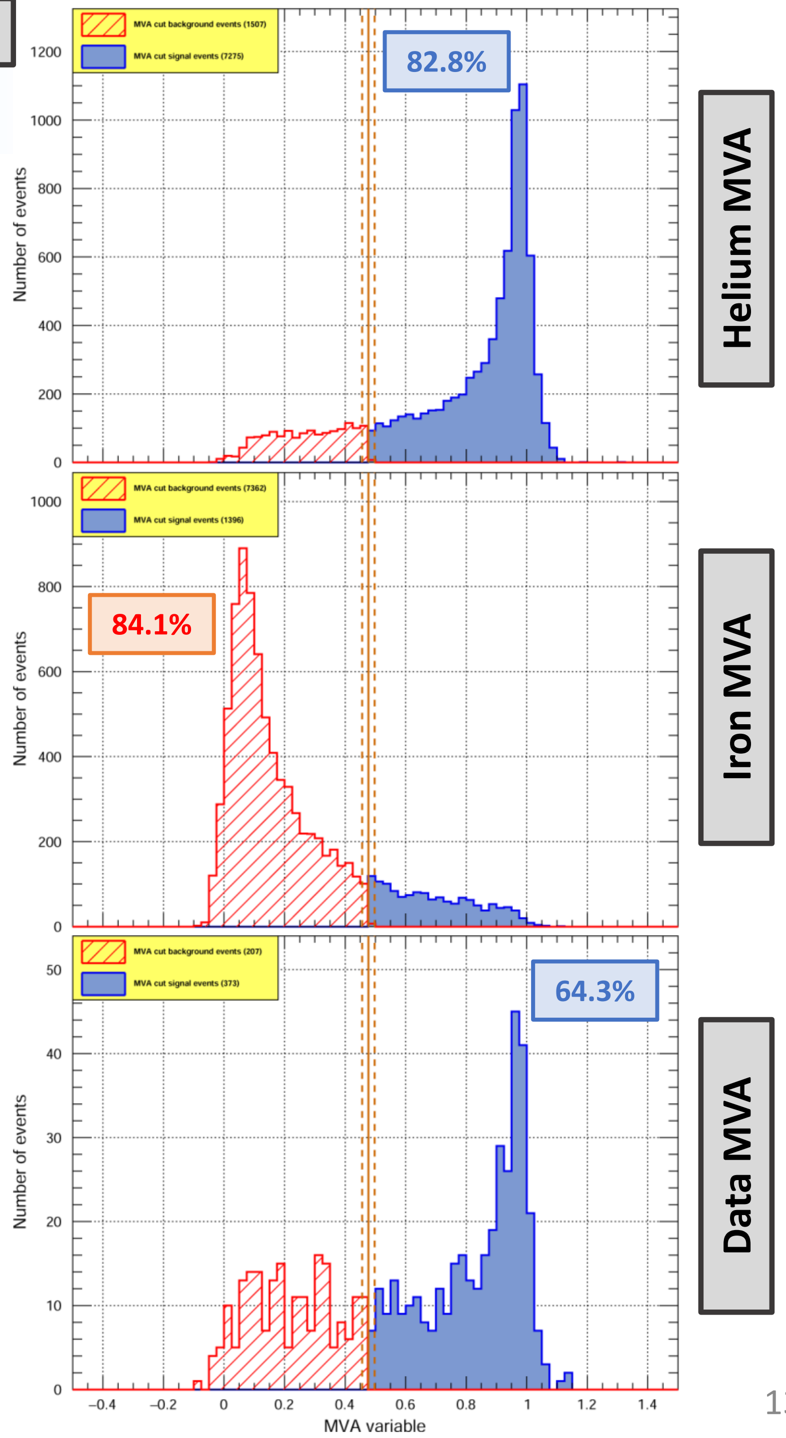




Preliminary analysis

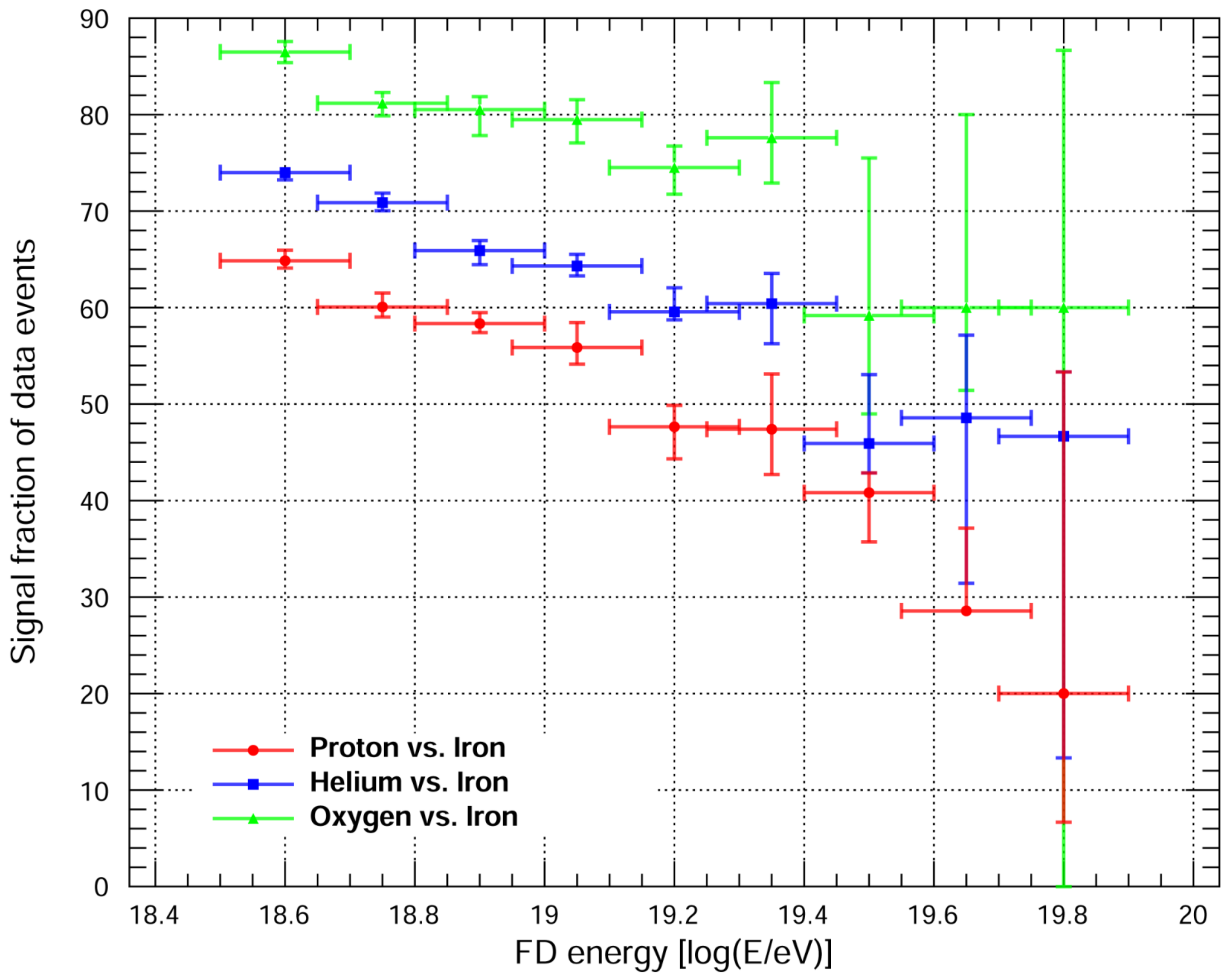


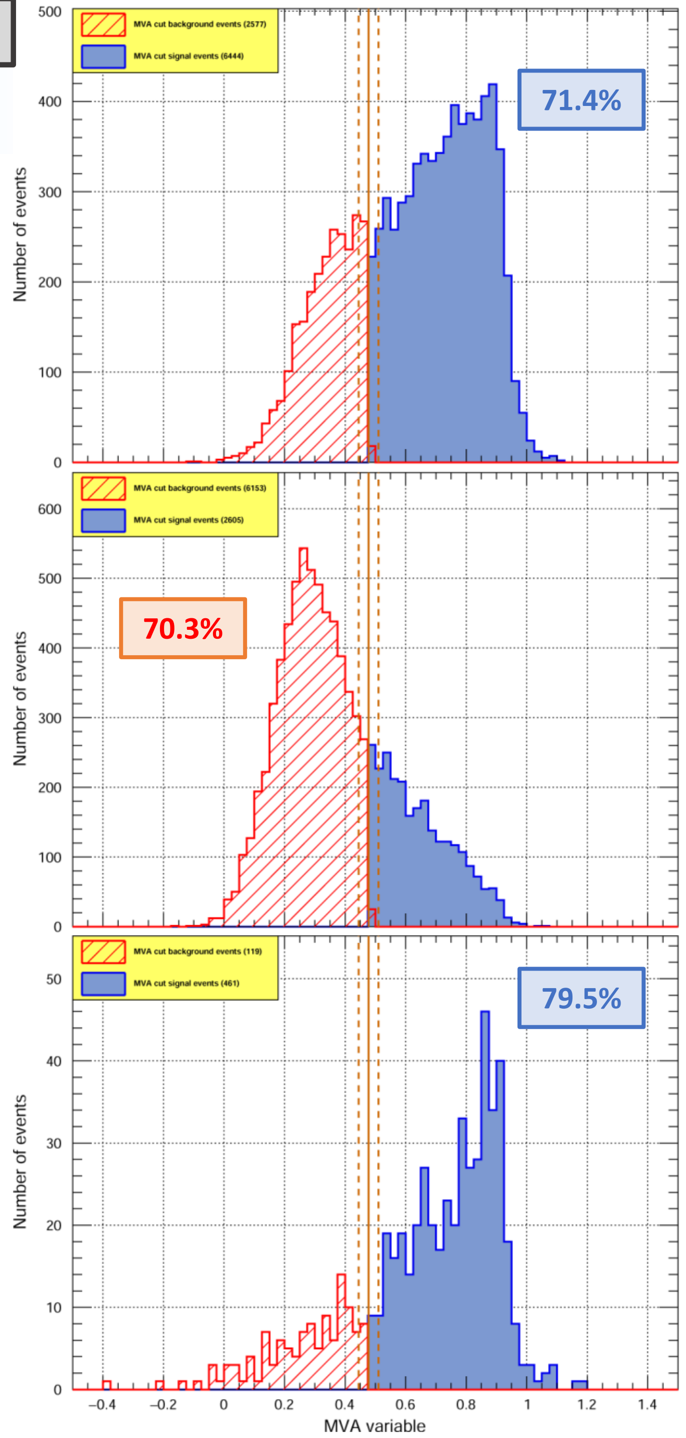
$10^{18.95} - 10^{19.15}$ eV



Preliminary analysis

$10^{18.95} - 10^{19.15}$ eV





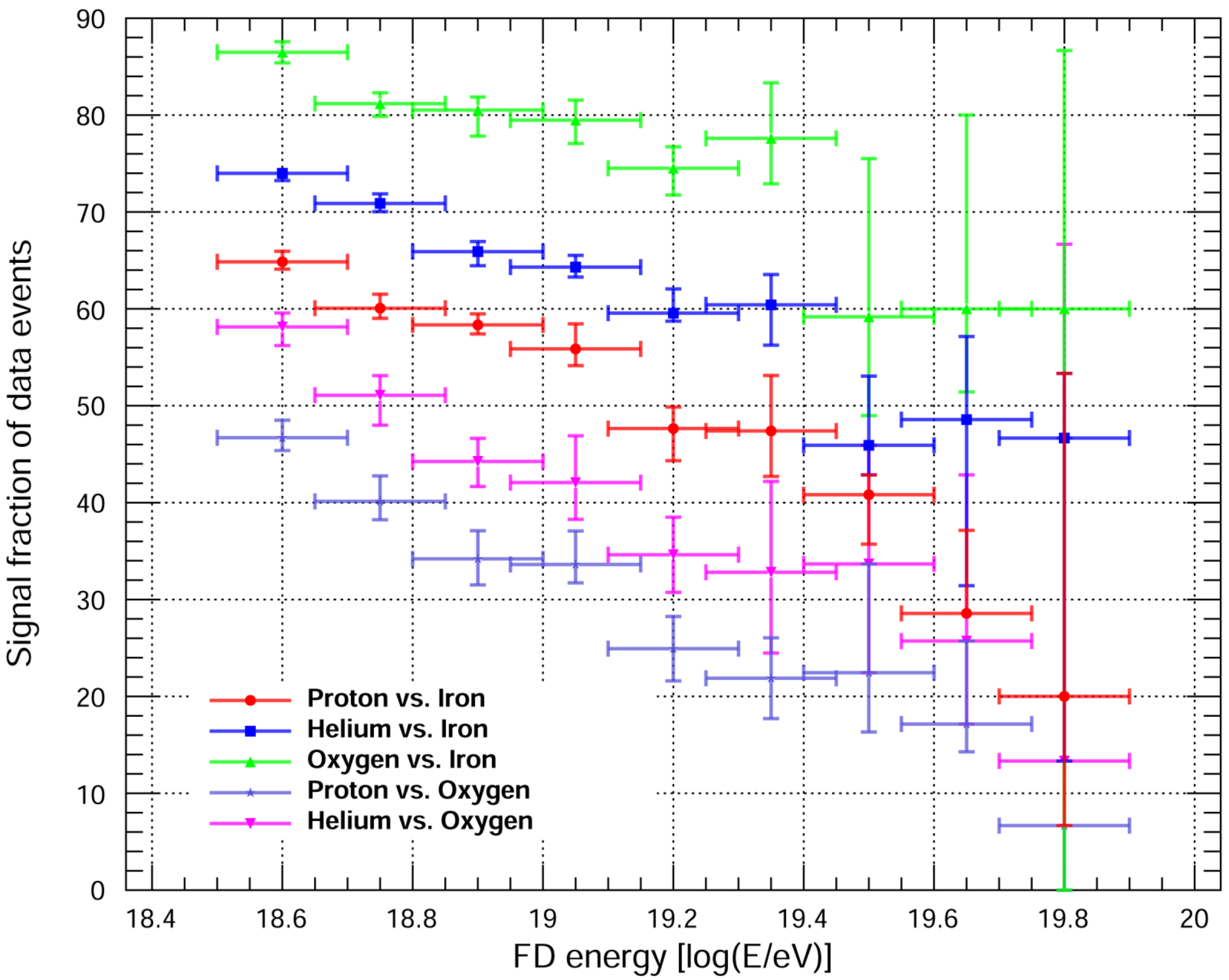
Preliminary analysis

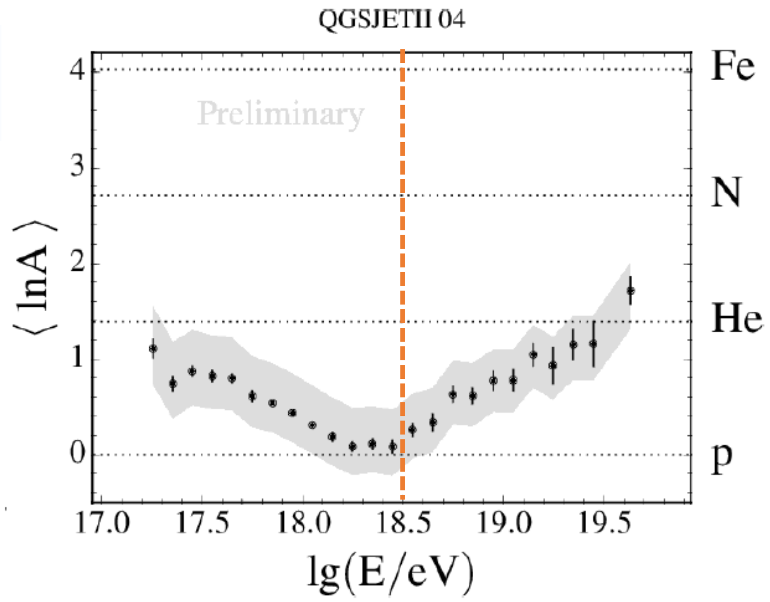
Multivariate analysis (MVA)

Preliminary analysis

Multivariate analysis (MVA)

Preliminary analysis





Mass composition becoming increasingly heavier with increasing energy

However!

Stability of analysis method still to be estimated

Summary and outlook

- Summary:
  - Analysis that includes many observables
  - Custom created analysis software (simple addition of new observables)
  - Separation between primary particle types from simulations
  - Possibility of event-by-event estimation of primary particle type
- Outlook:
  - Evaluation of uncertainties and study of analysis method performance
  - Perform analysis with other MVA methods and other high energy hadron interaction models
  - Include additional mass composition sensitive observables (Pierre Auger Observatory upgrade)

Backup slides

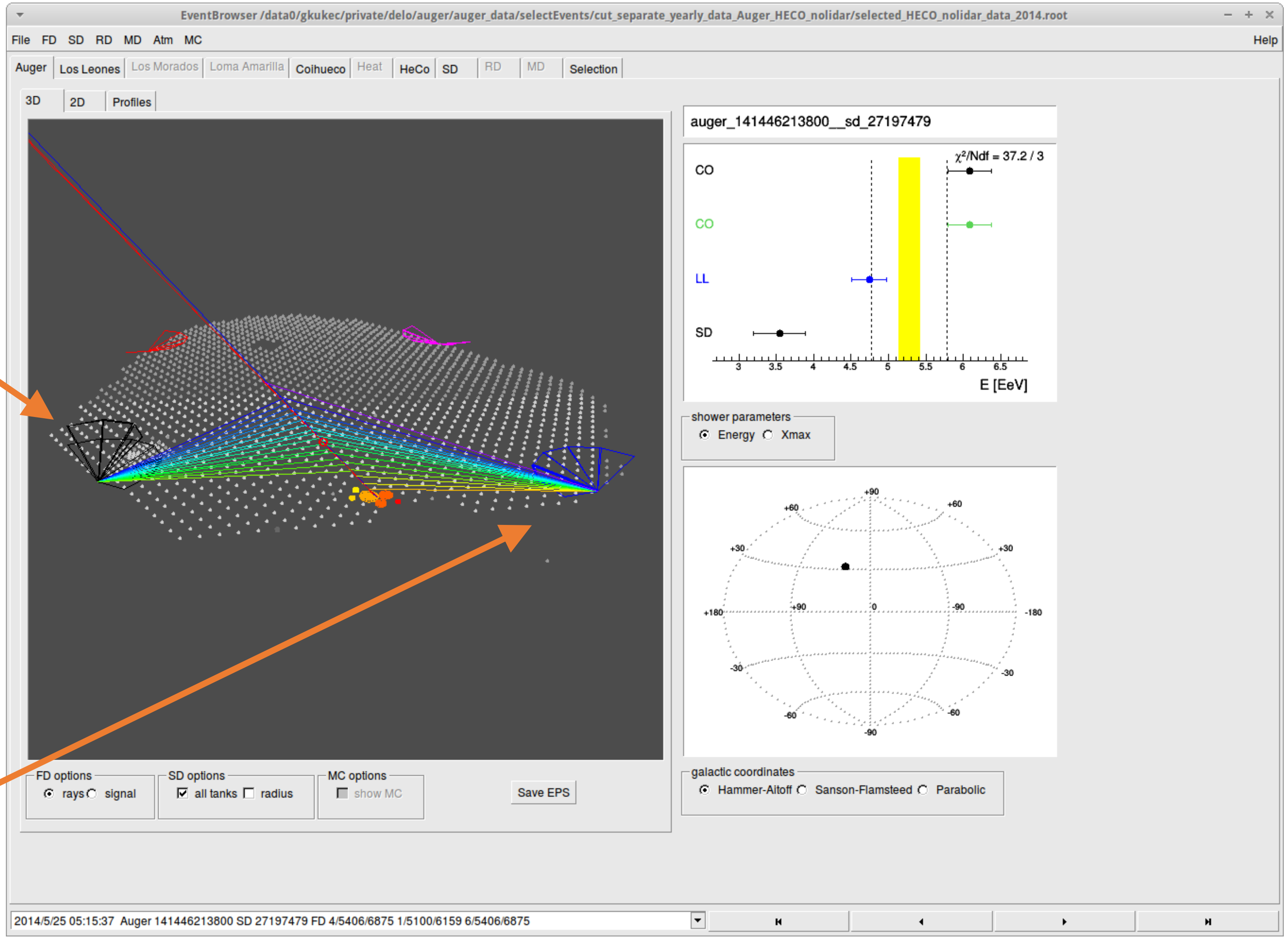
Analysis software

- All FD eyes rewritten from events in ADST format into a ROOT file with manageable size (quicker)
- Example event with two active eyes

Coihueco FD eye with energy $E = 6.09 \times 10^{18}$

If energy selection cut is set between the two, this event is taken into analysis twice

Los Leones FD eye with energy $E = 4.74 \times 10^{18}$



MVA cut uncertainty

- Uncertainty on MVA cut is automatically calculated from selected "data" tree
- Sets of observables are plotted to determine their correlation ρ_{ij}
- MVA analysis transforms input values according to the "signal" and "background" trees with:

$$X_{norm} = 2 \cdot \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Transforming all observables accordingly, we get their uncertainties σ_i
- Calculating covariance of two observables as $cov(i,j) = \rho_{ij}\sigma_i\sigma_j$
- The covariance matrix is diagonalised
- The final uncertainty on the MVA cut σ for N observables is then:

$$\sigma^2 = \sum_{i=1}^{N} cov(i,i)^2$$

MVA cut uncertainty

- Best for MVA are uncorrelated observables
- Uncertainty on MVA cut comes from individual observable uncertainties and is calculated from correlation matrix
- Systematic uncertainties due to MVA cut selection need to be estimated