Multivariate analysis results from the Telescope Array

Mikhail Kuznetsov, Grigory Rubtsov and Yana Zhezher

for the Telescope Array collaboration

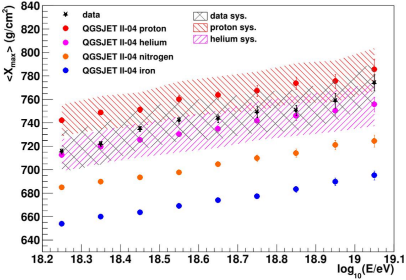
COSPA-2019
Brussels
October 4, 2019

Supported by Russian Science Foundation

- Motivation for multivariate analysis in cosmic-ray experiments
- Features of multivariate analysis in Telescope Array
- TA MVA results & prospects
  - UHECR mass composition with TA surface detector
  - Prospect for UHECR mass composition anisotropy with TA SD
  - Search for UHE photons with TA SD
  - Search for UHE neutrinos with TA SD

Problems of UHECR

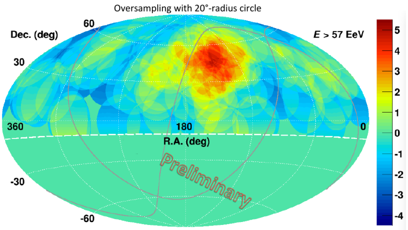
UHECR mass composition?



Standard approach:

- To study $X_{\max}$ distibution of showers

UHECR sources?

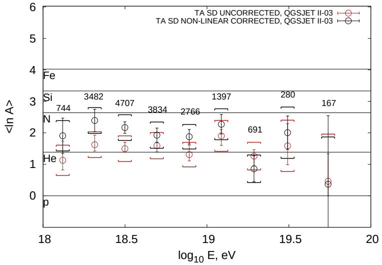


Standard approaches:

- To study correlation of events with sources catalogs
- To study cross-correlations of events
- To fit UHECR spectrum and/or composition with source models

Problems of UHECR

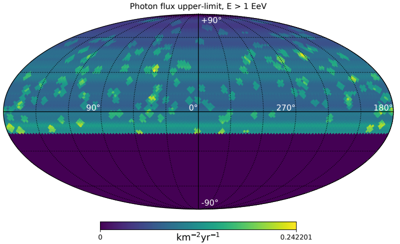
UHECR Mass composition?



Complementary approach:

- To study full imprints of showers in surface detector

UHECR sources?

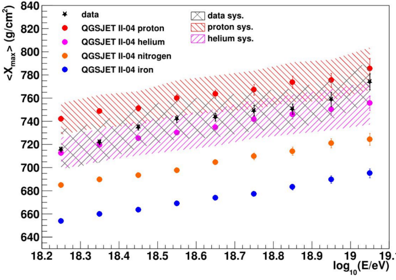


Complementary approaches:

- To search for UHE gamma signal
- To search for UHE neutrino signal

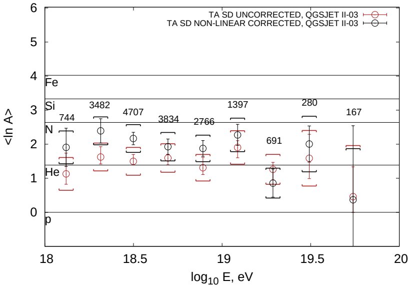
Multivariate analysis is needed!

Mass composition from FD



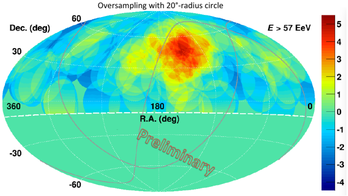
- Lower statistics
- Lower flexibility of analysis: only one observable - $X_{\max}$
  - Harder to go beyond $\langle \ln A \rangle$ analysis
- Lower dependency of hadronic interaction model

Mass composition from SD



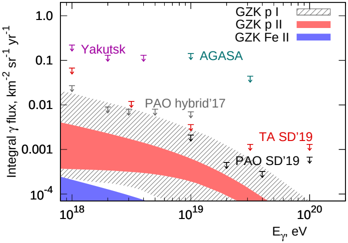
- Higher statistics (especially at highest energies)
  - A way to composition anisotropy study
- Higher flexibility of analysis: many observables
  - Easier to test multi-component models
- Higher dependency on hadronic interaction model
  - Possibly could be reduced with machine learning

Correlation analysis & sources models fits



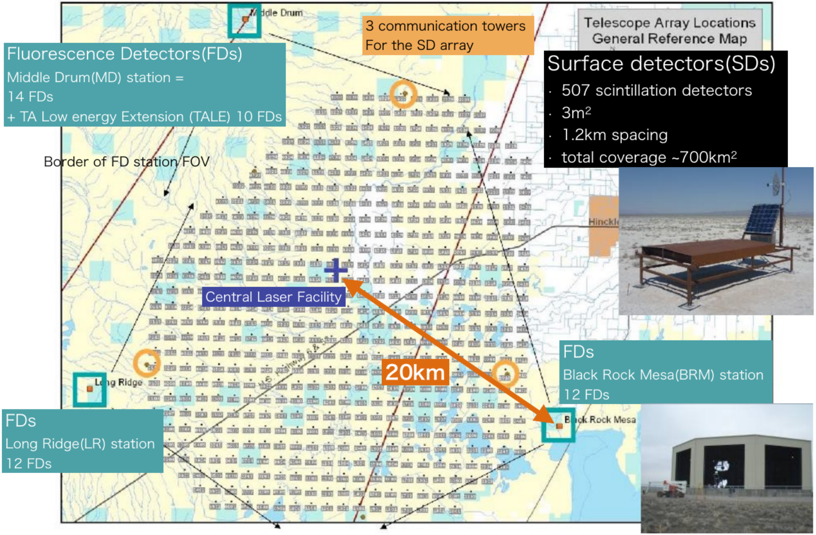
- Events arrival directions are smeared by magnetic fields
- High uncertainty in primary mass composition estimation (i.e. the strength of smearing is highly uncertain)
- Search for DM decay is difficult due to large background
- Higher statistics

Multimessenger analysis



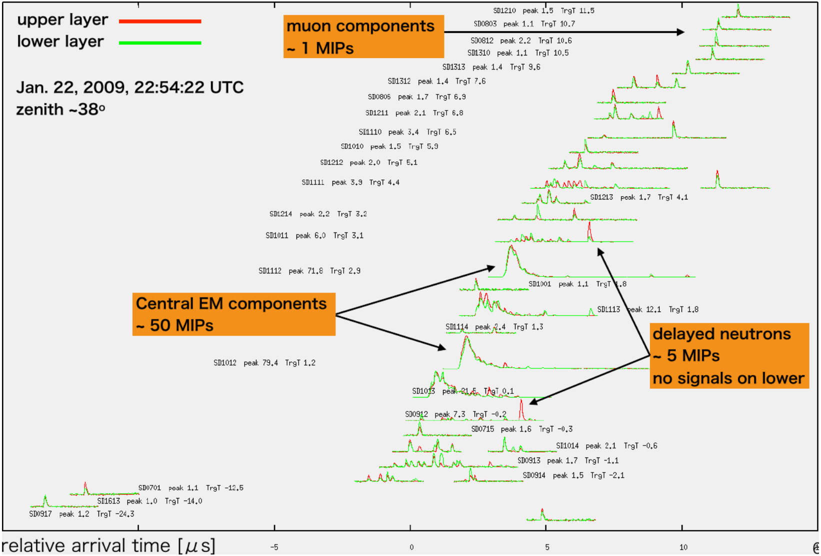
- Events arrival directions does not depend on magnetic fields
- Secondary γ and ν fluxes for various primaries differ significantly
  - A way to distinguish primaries (but uncertainty due to propagation effects)
- UHE γ flux is the most sensitive tool for DM decay search
- Lower statistics (no γ's or ν's were found up to now :)
  - The sensitivity could be increased with machine learning

Telescope Array experiment



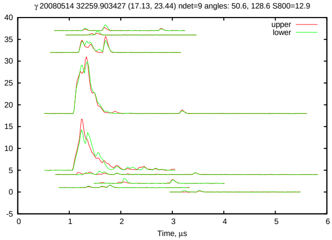
Largest UHECR statistics in the Northern Hemisphere

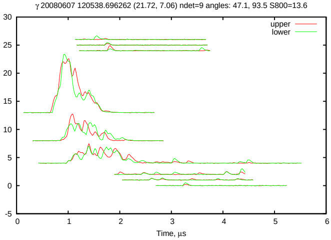
Sample of TA surface detector event



p-induced EAS

γ-induced EAS





Photon-induced showers:

- arrive younger
- contain less muons
- multiple SD observables affected: **front curvature, Area-over-peak, $\chi^2/d.o.f.$, etc.**

List of relevant SD observables for MVA

1. Zenith angle, θ;
2. Signal density at 800 m from the shower core, S_{800};
3. Linsley front curvature parameter, a;
4. Area-over-peak (AoP) of the signal at 1200 m;

Pierre Auger Collaboration, Phys.Rev.Lett. 100 (2008) 211101

5. AoP LDF slope parameter;
6. Number of detectors hit;
7. N. of detectors excluded from the fit of the shower front;
8. $\chi^2/d.o.f.$;
9. $S_b = \sum S_i \times r^b$ parameter for $b = 3$ and $b = 4.5$;

Ros, Supanitsky, Medina-Tanco et al. Astropart.Phys. 47 (2013) 10

10. The sum of signals of all detectors of the event;
11. Asymmetry of signal at upper and lower layers of detectors;
12. Total n. of peaks within all FADC traces;
13. N. of peaks for the detector with the largest signal;
14. N. of peaks present in the upper layer and not in lower;
15. N. of peaks present in the lower layer and not in upper;

Multivariate analysis

How to deal with this large amount of observables?

Machine learning for multivariate analysis.

- The Boosted Decision Trees (BDT) technique is used to build event classifier based on multiple observables.
- root::TMVA is used as a stable implementation.

  PoS ACAT 040 (2007), arXiv:physics/0703039

- BDT is trained with two Monte-Carlo sets: signal and background*
- BDT classifier is used to convert the set of observables of each event to a number $\xi \in [-1 : 1]$
- ξ is available for one-dimensional analysis.

* Depending on a task the signal could be either γ, ν or *Fe* events and the background is *p* events.

Multivariate analysis

How to deal with this large amount of observables?

Machine learning for multivariate analysis.

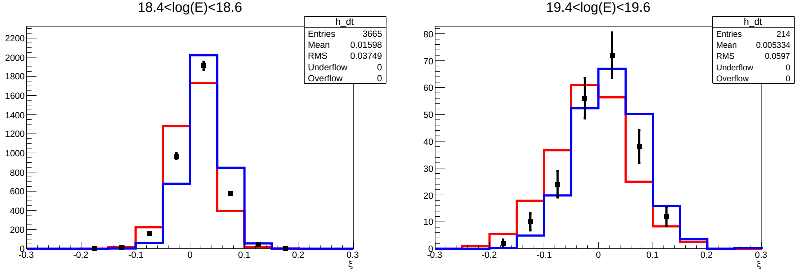
- The Boosted Decision Trees (BDT) technique is used to build event classifier based on multiple observables.
- root::TMVA is used as a stable implementation.

PoS ACAT 040 (2007), arXiv:physics/0703039

- BDT is trained with two Monte-Carlo sets: signal and background*
- BDT classifier is used to convert the set of observables of each event to a number $\xi \in [-1 : 1]$
- ξ is available for one-dimensional analysis.

* Depending on a task the signal could be either γ, ν or *Fe* events and the background is *p* events.

TA MVA Composition analysis

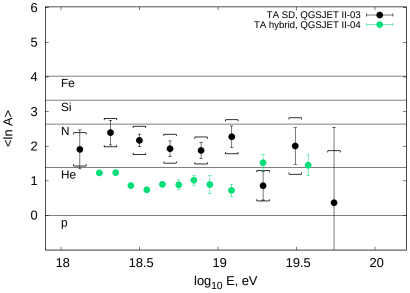


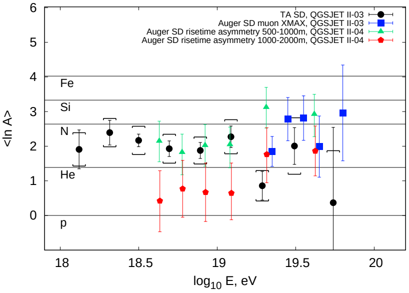
data *p* MC *Fe* MC

- Two-component approximation is used, first approximation of atomic mass is
  $\langle \ln A \rangle^{(1)} = \epsilon_p \times \ln(M_p) + \epsilon_{Fe} \times \ln(M_{Fe})$
- root::TFractionFitter is used for binned template fitting.

Telescope Array, PRD 99, 022002 (2019)

TA MVA Composition analysis: results



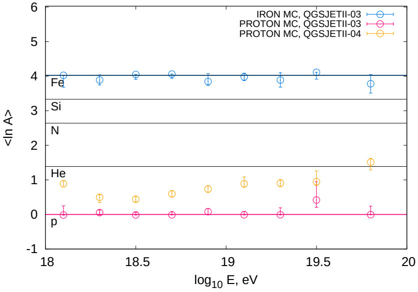


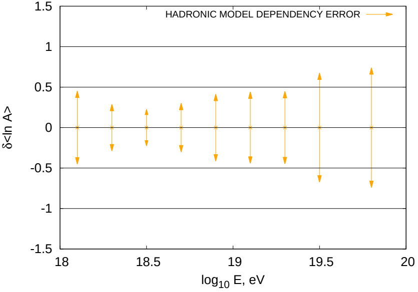
$$\langle \ln A \rangle = 2.0 \pm 0.1(stat.) \pm 0.44(syst.)$$

Consistent with Auger SD results, qualitatively consistent with TA FD given hadronic model systematics and systematics of FD

TA MVA composition analysis: hadronic model systematic

Comparison between QGSJet II-03 and QGSJet II-04 models:





$$\delta \ln A_{hadr.} = 0.4$$

Possibly solvable with more hadronic models input and advanced machine learning techniques

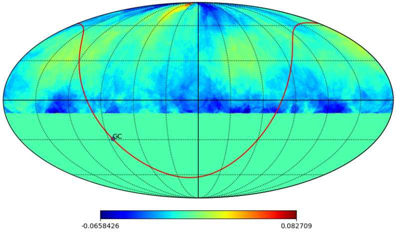
TA MVA composition analysis: prospect for anisotropy

Test of sensitivity with MC sets

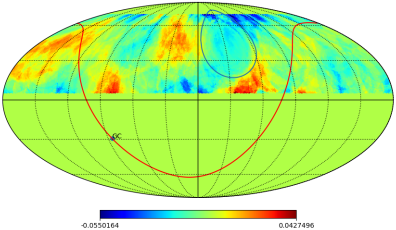
Telescope Array, PoS(ICRC2019)494

- Isotropic set with data composition (p + Fe mixture).
- Set with data composition and light "hotspot" at energies $\log E > 19.2$.
- Set with data composition and heavy "hotspot" at energies $\log E > 19.2$.

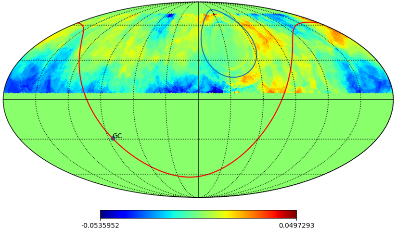
Isotropic composition



Light "hotspot"



Heavy "hotspot"



Anisotropy test: HEALPix maps comparison

How to take into account large-scale gradients of ξ ?

- Spatial distribution of average ξ may be pixelized: HEALPix

  Gorski et. al., Astrophys.J, 622, 759, 2005
- HEALPix transforms skymap into one-dimensional array of pixels
- It is suitable to perform χ^2-test between two maps
- We split each skymap into 786 pixels (mean resolution is 7.3°) to have enough statistics in each pixel

Results

- Light "hotspot" vs. isotopic composition:
  $\chi^2/d.o.f. = 1.47$, $p = 6.9 \times 10^{-5}$ $\Rightarrow$ 3.8σ
- Heavy "hotspot" vs. isotopic composition:
  $\chi^2/d.o.f. = 1.63$, $p = 6.3 \times 10^{-7}$ $\Rightarrow$ 4.8σ

Anisotropy test: HEALPix maps comparison

How to take into account large-scale gradients of ξ ?

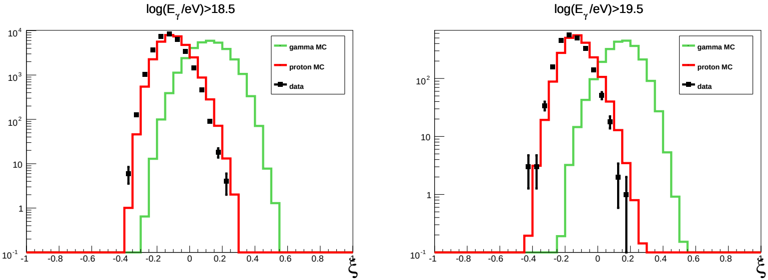
- Spatial distribution of average ξ may be pixelized: HEALPix

  Gorski et. al., Astrophys.J, 622, 759, 2005
- HEALPix transforms skymap into one-dimensional array of pixels
- It is suitable to perform χ^2-test between two maps
- We split each skymap into 786 pixels (mean resolution is 7.3°) to have enough statistics in each pixel

Results

- Light “hotspot” vs. isotopic composition:
  $\chi^2/d.o.f. = 1.47$, $p = 6.9 \times 10^{-5}$ $\Rightarrow$ 3.8σ
- Heavy “hotspot” vs. isotopic composition:
  $\chi^2/d.o.f. = 1.63$, $p = 6.3 \times 10^{-7}$ $\Rightarrow$ 4.8σ

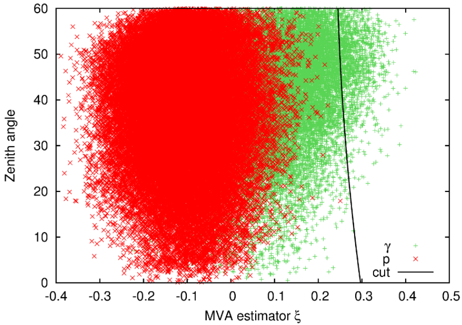
TA MVA Search for diffuse UHE photons



data photon MC proton MC

- The photon candidates are selected using the cut on ξ: $\xi > \xi_{cut}(\theta)$
- Cut is optimized in each energy range using proton and photon Monte-Carlo
- The null-hypothesis is assumed for the optimization (all events are protons)

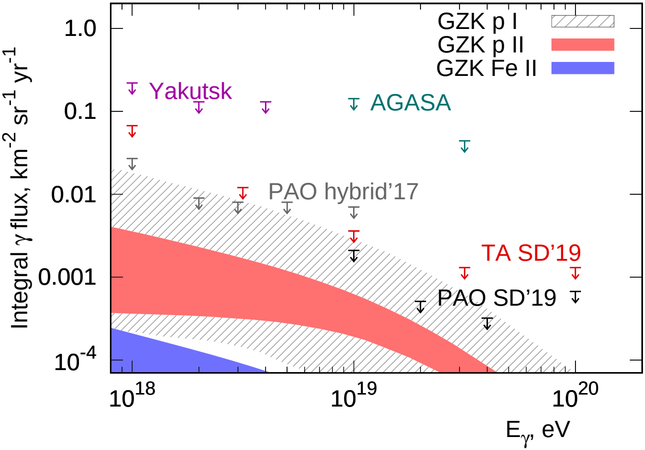
TA MVA Search for diffuse UHE photons



Effective exposure is estimated with γ MC assuming E^{-2} primary spectrum

E_γ	quality cuts	ξ-cut	A_{eff} km^2 sr yr
$10^{18.0}$	6.5%	9.8%	**77**
$10^{18.5}$	19.9%	10.6%	**255**
$10^{19.0}$	43.6%	16.2%	**852**
$10^{19.5}$	52.0%	37.2%	**2351**
$10^{20.0}$	64.2%	52.3%	**4055**

Results: diffuse UHE photons flux limits



models from J. Alvarez-Muniz et al. EPJ Web Cong. ***53****, 01009 (2013)*

$E_\gamma >$, eV	$10^{18.0}$	$10^{18.5}$	$10^{19.0}$	$10^{19.5}$	$10^{20.0}$
γ candidates	1	0	0	0	1
$F_\gamma <$	0.067	0.012	0.0036	0.0013	0.0013

TA MVA search for point sources of UHE photons

- Independent search for γ in each skymap direction
- The angular size of the each search region is equal to the γ **angular resolution:**

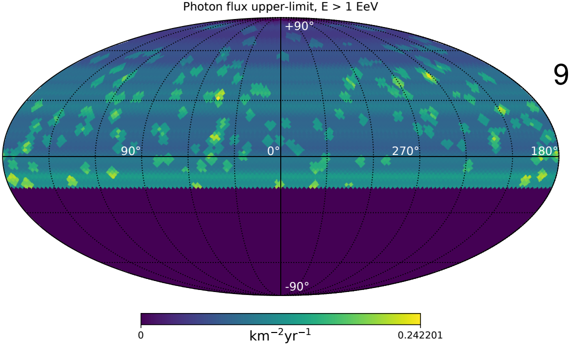
$E_\gamma \geq$, eV	$10^{18.0}$	$10^{18.5}$	$10^{19.0}$	$10^{19.5}$	$10^{20.0}$
ang.res.	3.00°	2.92°	2.64°	2.21°	2.06°

- The skymap is pixelized into 12288 directions with HEALpix (7868 in TA field of view)

Optimisation of MVA-cut for γ flux upper-limit:

- Assume the flux consists of protons only (null hypothesis): $F_{\rm total} = F_{\rm p}$
- Optimize the ξ-cut separately for the best upper-limit in each direction using MC p and MC γ
- E^{-2} γ-spectrum is assumed

Results: point-source photon flux upper-limits



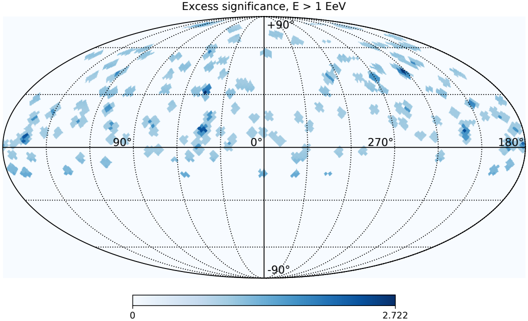
95% C.L.

$E_\gamma \geq$, eV	$\langle F_\gamma \rangle \leq$, $\mathrm{km}^{-2}\mathrm{yr}^{-1}$
$10^{18.0}$	0.094
$10^{18.5}$	0.029
$10^{19.0}$	0.010
$10^{19.5}$	0.0071
$10^{20.0}$	0.0058

Pierre Auger: $\langle F_\gamma \rangle \leq 0.035\ \mathrm{km}^{-2}\mathrm{yr}^{-1}$ (1° ang.res., $10^{17.3} \leq E \leq 10^{18.5}$ eV)

A. Aab et al. ApJ 789, 160 (2014)

Results: photon excesses significance



$E_\gamma \geq$, eV	max. γ signif. (pre-trial)
$10^{18.0}$	2.72 σ
$10^{18.5}$	2.71 σ
$10^{19.0}$	2.89 σ
$10^{19.5}$	2.76 σ
$10^{20.0}$	3.43 σ

The excesses are insignificant, given approx. 1000 of independent trials

Target search for photons from dwarf galaxies

Probe for the possible decay of heavy dark matter (HDM)

- HDM decay produce significant amount of photons in any model
  M. Kachelriess et al., PRD 98, 083016 (2018)
- DM is abundant in dwarf galaxies (Galactic Center is outside the TA field of view)
- Target source set: 21 dwarf galaxies — satellites of Milky Way
  V. Bonnivard et al., MNRAS 453 (2015), 849
- Search for γ in stacked skymap directions of dwarf galaxies (pixel size = γ **ang.res.**)

Results

No evidence for photon signal ($N_\gamma^{\mathrm{cand.}} = 0$ at all energies)

E_γ, eV	$10^{18.0}$	$10^{18.5}$	$10^{19.0}$	$10^{19.5}$	$10^{20.0}$
$F_{UL}^\gamma, \mathrm{km}^{-2}\mathrm{yr}^{-1}$	0.15	0.057	0.014	0.0076	0.0052

These results can be used to constrain HDM models

Target search for photons from dwarf galaxies

Probe for the possible decay of heavy dark matter (HDM)

- HDM decay produce significant amount of photons in any model
  M. Kachelriess et al., PRD 98, 083016 (2018)
- DM is abundant in dwarf galaxies (Galactic Center is outside the TA field of view)
- Target source set: 21 dwarf galaxies — satellites of Milky Way
  V. Bonnivard et al., MNRAS 453 (2015), 849
- Search for γ in stacked skymap directions of dwarf galaxies (pixel size = γ **ang.res.**)

Results

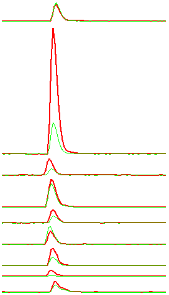
No evidence for photon signal ($N_\gamma^{\mathrm{cand.}} = 0$ at all energies)

E_γ, eV	$10^{18.0}$	$10^{18.5}$	$10^{19.0}$	$10^{19.5}$	$10^{20.0}$
$F^\gamma_{UL}, \mathrm{km}^{-2}\mathrm{yr}^{-1}$	0.15	0.057	0.014	0.0076	0.0052

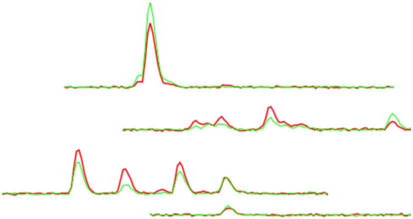
These results can be used to constrain HDM models

Neutrino search strategy

proton shower, 78.3°

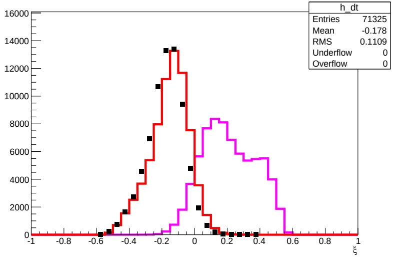


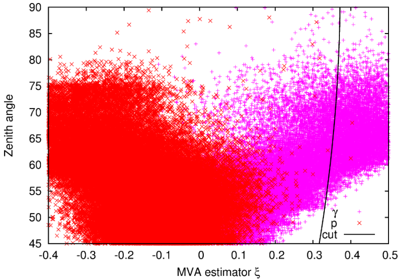
neutrino shower, $\theta = 78.6°$



- Only inclined events are studies 45° $< \theta \ <$ 90°
- No energy cut
- The optimization of ξ-cut is similar to the photon search case

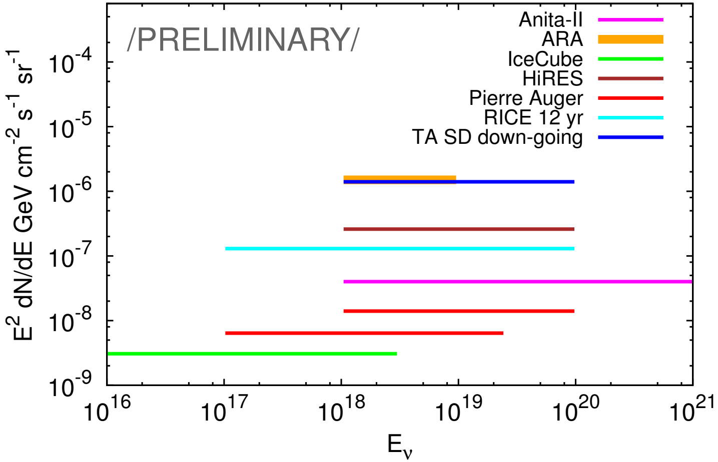
Distribution of MVA estimator (ξ) for data and MC





- Geometric exposure for $\theta \in (45^\circ, 90^\circ)$: 8042 $\mathrm{km}^2\,\mathrm{sr}\,\mathrm{yr}$
- probability to interact in the atmosphere: 1.4×10^{-5}
- trigger, reconstruction and quality cuts efficiency $\sim 7\%$
- ξ cut efficiency: $\sim 24\%$
- total exposure (all flavors): $A = 1.9 \times 10^{-3}\,\mathrm{km}^2\,\mathrm{sr}\,\mathrm{yr}$

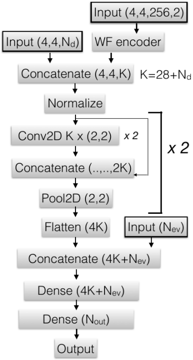
Results

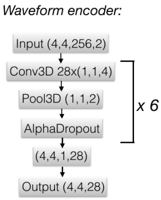


- 0 neutrino candidates after cuts
- Single flavor diffuse neutrino flux limit for $E > 10^{18}$ eV: $E^2 f_\nu < 1.4 \times 10^{-6}\,\mathrm{GeV\,cm^{-2}s^{-1}sr^{-1}}$ (90% C.L.)

Plans

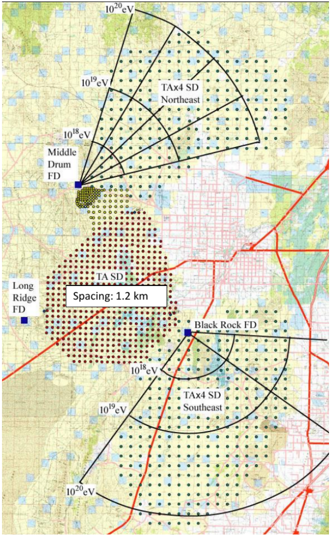
To improve MVA sensitivity with recently developing convolutional Neural Net





Telescope Array, PoS(ICRC2019)30

To increase statistics with recently deploying TAx4 SD detector



Conclusions

Multivariate analysis + machine learning is a powerful tool for UHECR study

- UHECR mass composition without X_{max}
- UHECR composition anisotropy
- Sensitive UHE photon and neutrino search

Thank you!

Conclusions

Multivariate analysis + machine learning is a powerful tool for UHECR study

- UHECR mass composition without X_{max}
- UHECR composition anisotropy
- Sensitive UHE photon and neutrino search

Thank you!