
Oleg Kalashev
INR RAS
Moscow

ACAT 2019
March 11, 2019  Saas Fee
Cosmic rays

Different Physics in each energy range:

Solar modulation: 
\[ 10^8 \text{eV} \leq E \leq 10^{11} \text{eV} \]

Galactic sources: 
\[ 10^{11} \text{eV} \leq E \leq 10^{18} \text{eV} \]

Extragalactic sources: 
\[ E > 10^{18} \text{eV} \]

GZK cut-off 
\[ E \simeq 10^{20} \text{eV} \]

Problem:
Extremely small flux, hard to observe directly
Questions

• Sources (extragalactic)

• Production mechanism

Observables (indirect)

• Energy spectrum

• Mass composition

• Arrival directions

Observables (direct)

• EAS properties observable from Earth (density profile on SD, fluorescence light on FD)
Questions

• Sources (extragalactic)
• Production mechanism

Observables (indirect)

• Energy spectrum
• Mass composition
• Arrival directions

Observables (direct)

• EAS properties observable from Earth (density profile on SD, fluorescence light on FD)

B and R must be big enough to keep particles inside in the acceleration region for enough time.
Questions

- Sources (extragalactic)
- Production mechanism

Observables (indirect)

- Energy spectrum
- Mass composition
- Arrival directions

Observables (direct)

- EAS properties observable from Earth (density profile on SD, fluorescence light on FD)
UHECR Detection Methods

**Surface detectors:**
Duty cycle ~ 95%

**Fluorescence detectors:**
Duty cycle ~ 10%
Modern UHECR experiments

- Southern hemisphere: Pierre Auger Observatory (Argentina)
  - 1660 Cherenkov detectors (water tanks), placed at distance 1.5 km from each other (hexagonal grid) covering the surface above 3000 km$^2$
  - 4 fluorescence telescope stations in the corners
  - International collaboration (16 countries)
Modern UHECR experiments

- Northern hemisphere: Telescope Array (Utah, USA).
Auger and TA spectra

Full Sky spectra

Common declination band (-15.9° < δ < 24.8°)

Better agreement between Auger and TA in the common declination band

Declination dependence?
Composition of UHECR

EAS produced by different mass nuclei are very similar - hard to distinguish at the level of individual events

- E=1-few EeV: light nuclei
- Heavier composition at larger energies

Deflection angles in Milky Way > 20 deg.
Cosmic rays may not point to the source

Auger 2017
Arrival directions and their interpretation

- Accuracy for nuclei primaries: 1.5 deg, gamma primaries: 2-3 deg

- Interpretation for nuclei depends on the assumption on the composition. Extra information needed.

- Multimessenger approach (study secondary signals from CR interactions)

  - GZK photons and neutrinos may point back to source (expected energies $E \sim E_{\text{EeV}}$) Not observed yet.

  **Sensitivity to diffuse and directional flux should be enhanced.**
Ways to improve gamma sensitivity.

• For diffuse flux: Improve discrimination of gamma and nuclei induced showers

• For directional flux search: enhance angular resolution for both nuclei and gamma to reduce the background
Ways to improve gamma sensitivity.

- For diffuse flux: Improve discrimination of gamma and nuclei induced showers

  Construct the function of observables most sensitive to composition using NN *in progress*

- For directional flux search: enhance angular resolution for both nuclei and gamma to reduce the background

*this talk*
Sample event

zenith ~38°

Time step 20 ns

relative arrival time [μs]
Event reconstruction

standard parametric approach

• LDF

\[ f(r) = \left( \frac{r}{R_m} \right)^{-1.2} \left( 1 + \frac{r}{R_m} \right)^{-(\eta - 1.2)} \left( 1 + \frac{r^2}{R_1^2} \right)^{-0.6} \]

\[ R_m = 90.0 \text{ m}, \quad R_1 = 1000 \text{ m}, \quad R_L = 30 \text{ m}, \quad \eta = 3.97 - 1.79 (\sec (\theta) - 1), \]
\[ r = \sqrt{(x_{\text{core}} - x)^2 + (y_{\text{core}} - y)^2}, \]

• Timing

\[ t_r = t_o + t_{\text{plane}} + a \times (1 + r/R_L)^{1.5} LDF (r)^{-0.5} \]

\[ LDF (r) = f(r) / f(800 \text{ m}) \quad S(r) = S_{800} \times LDF (r) \]

Free parameters:
\[ x_{\text{core}}, y_{\text{core}}, \theta, \phi, S_{800}, t_0, a \]

Observables:
\[ t_r \quad - \text{detector time} \]
\[ S_r \quad - \text{detector integral signal} \]
Event reconstruction

standard parametric approach

Energy estimate

\[ E'_{SD} = E'_{SD}(S800, \theta) \]

- table function
Event reconstruction

*Machine learning approach*

*Purpose (ideally):* recover primary particle properties (arrival direction, energy, mass, …) as function of observables.

**Direct observables in SD:**
- Time series of the SD signals

**Instruments:**
- SD Monte-Carlo (EAS development and detector response)
- Artificial neural network (NN)
  - Can describe any continuous function of input data
  - Can be tuned using examples generated using Monte-Carlo
Event reconstruction

Machine learning approach

*Purpose (ideally):* recover primary particle properties (arrival direction, energy, mass, …) as function of observables.

*In real life:*
- observables depend on unknown/random factors
- NN function defines optimal test statistic
- obtain corrections to parametric reconstruction

**Observables in SD:**
- Time series of the SD signals

**Instruments:**
- SD Monte-Carlo (EAS development and detector response)
- Artificial neural network (NN)
  - Can describe any continuous function of input data
  - Can be tuned using examples generated using Monte-Carlo
Method in nutshell

- Extract useful detector features using 1-D convolutions
- Treat detector network as a multichannel image using 2D convolution layers

First proposed:
Applied to Pierre Auger geometry with toy Monte Carlo

Track 2 talk
SD reconstruction NN architecture

Dimensions: 
(N, N, T, 2)

Waveform
detector
layers

N=4-8, T=128-256

Standard SD reconstruction is used to center image around shower core
SD reconstruction NN architecture

Using 1D + 2D convolutions

Input \((4,4,256,2)\)

1-3 x Conv3D \(16 \times (1,1,4)\)

Pool3D \((1,1,4)\)

(4,4,1,16)

Reshape \((4,4,16)\)

1-3 x Conv2D

Pool2D

(1,1,256)

Flatten \((256)\)

Dense

Output

feature vectors for each detector
SD reconstruction NN architecture

Using 1D + 2D convolutions

Adding extra detector features:

- Exact detector position
- Detector state (on/off/saturated)
- Standard reconstruction parameters (integral signal, timing relative to plane front)
SD reconstruction NN architecture

Using 1D + 2D convolutions

Adding extra event features:

- season and time
- optionally standard reconstruction data (e.g. $S_{800}$)
SD reconstruction NN architecture

Using 1D + 2D convolutions

Simplifying task:

Modifying cost function - we calculate correction to standard reconstruction for angles and energy
SD reconstruction NN architecture

Using 1D + 2D convolutions

Further optimisation (optional):

- Use dropout regularization
- Replace heavy 2D-convolutions with depth-wise separable convolutions
- Using residual blocks (shortcuts)
Training the model

• Minimizing mean square error
• Adaptive learning rate (adadelta optimizer arxiv 1212.5701)
• Number of training samples $\sim 10^6$ (100 GB data) - do not fit into RAM). hdf container is used и generator API in keras
• Number of weights to learn $10^5 - 10^6$
• Regularization to avoid overfitting:
  • L2
  • dropout
  • noise layers
• Optimizing network architecture hyper-parameters (hyperopt package)
• Hardware: NVIDIA GTX-1080-Ti GPU
• Instruments: python, numpy, tensorflow, keras, h5py
EAS modelling

- **MC**: CORSIKA
- **HE** hadronic interactions: QGSJETII-03 (QGSJETII-04 in preparation)
- **LE** hadronic interactions: FLUKA
- **EM processes**: EGS4
- **Detector response**: GEANT4
- **Event sampling**:
  - Energy sampling $E^{-1}$
  - Mass composition: H, He, N, Fe (1:1:1:1)
  - Isotropic primary flux with zenith angles $< 45$ degrees
  - Standard energy spectrum reconstruction cuts applied
How to see that model does job
in presence of unavoidable uncertainty

Explained variance score

$$EV(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

$y$ - true value of quantity being predicted (in our case, error of parametric reconstruction)

$\hat{y}$ - model estimate of $y$
How to see that model does job in presence of unavoidable uncertainty

Explained variance score

\[ EV(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)} \]

\( y \) - true value of quantity being predicted (in our case, error of standard reconstruction)

\( \hat{y} \) - model estimate of \( y \)

More visually:
Compare error distribution in two approximations
Preliminary results

Zenith angle reconstruction errors

$\Delta \cos(\theta)$

Logarithmic scale for $E/\text{EeV}$

**nuclei**

Logarithmic scale for $E/\text{EeV}$

**photons**

Logarithmic scale for $E/\text{EeV}$

Graphs showing the distribution of $\Delta \cos(\theta)$ for different energy ranges.
Preliminary results

Energy reconstruction errors (nuclei primaries)

\[ \Delta E / E \]

\[ \log(E / EeV) \]
Conclusion

• NN allows to enhance substantially the accuracy of geometry and energy reconstruction
• Future plans:
  • Build photon-nucleon classifier
  • Investigate influence of cuts, enhance exposure
  • Study impact of systematics (hadronic interaction model, etc.)
  • Try more network architectures
  • Mass composition study
Appendix
Surface Detectors

- 3m² x 1.2cm x 2 layers
- WLSF: φ1mm 2cm spacing
- PMT for each layer

- 12bit 50MHz FADC x 2 layers
- CPU: Renesas SH4(25MHz)
- GPS, WLAN-modem
- Charge controller
SD energy from the hadronic models relative to the FD

SD energy from the hadronic models after normalization at $10^{19}$ eV
Hotspot with 9yr data (>57EeV)

Matthews, ICRC2017 TA highlight

With original 20° oversampling, spot looks larger ... Thus, scan over 15°, 20°, 25°, 30° and 35°

With 25° oversampling, significance maximum ~3 σ

<table>
<thead>
<tr>
<th>Binsize</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local</td>
<td>Global</td>
<td>Local</td>
<td>Global</td>
<td>Local</td>
</tr>
<tr>
<td>Year 5</td>
<td>5.12</td>
<td>3.14</td>
<td>5.43</td>
<td>3.55</td>
<td>5.16</td>
</tr>
<tr>
<td>Year 7</td>
<td>4.92</td>
<td>2.84</td>
<td>5.37</td>
<td>3.44</td>
<td>5.65</td>
</tr>
<tr>
<td>Year 9</td>
<td>4.42</td>
<td>2.06</td>
<td>4.72</td>
<td>2.50</td>
<td>5.06</td>
</tr>
</tbody>
</table>
SD reconstruction NN architecture

Problem: how to better take absent/not functioning detectors

• The event may occur close to detector network boundary
• Part of detectors may be turned off
Problem: how to better take absent/not functioning detectors

- The event may occur close to detector network boundary
- Part of detectors may be turned off

Dropout, the regularisation method in NN, simulate this situation
Dealing with absent/not working detectors:

Dropout, the regularisation method in NN

Srivastava et. al JMLR 15 (2014)

- In training mode neurons are switched off with probability $p$
- For $p=0.5$ we train simultaneously $2^n$ thinned neural networks
- In prediction mode neurons are on but their output weights are multiplied by $p$ (we average predictions of thinned nets)
SD reconstruction NN architecture

Take into account absent/not working detectors

Modified dropout

- Weights are corrected using fraction of working detectors
- In training mode part of the detectors may be switched off as in conventional dropout method
SD reconstruction NN architecture

Naive variant: на using 3D - convolution

Problem: incomparable scales in $t$ and $L$

- $L$ unit: 1200 meters
- $t$ unit: 20ns ~ 6 meters

Adjacent detectors waveforms are weakly correlated
Cuts applied on MC samples.

1. Each event must include at least 5 counters.
2. The reconstructed primary zenith angle must be less than 45°.
3. The reconstructed event core must be more than 1200 m from edge of the array.
4. Both the timing and lateral distribution fits must have $\chi^2$/degree of freedom value less than 4.
5. The angular uncertainty estimated by the timing fit must be less than 5°.
6. The fractional uncertainty in S(800) estimated by the lateral distribution fit must be less than 25%.
Neural networks

- Perceptron
- Multilayer perceptron networks
- Overfitting and regularization
- Convolutional neural networks
Perceptron

Combination of perceptrons can be used to build any logical operation

How to find proper weights:

- replace step function with continuous approximation
- adjust weights with gradient decent

\[ a = \sigma \left( \sum_{k} w_j x_j + b \right) \]

\[ \sigma(z) \equiv \frac{1}{1 + e^{-z}} \]
Multilayer perceptron (MLP)

\[ a_j^l = \sigma \left( \sum_k w_{jk}^l a_{k}^{l-1} + b_j^l \right) \]

**Theorem:** any finite continuous function can be approximated with any given accuracy by MLP

**Learning by minimising loss function:**

\[ C(w, b) \equiv \frac{1}{2n} \sum_x \| y(x) - a \|^2 \]

\( y(x) \) - network output  
\( a \) - desired output

**Back-propagation algorithm:** calculate all derivatives in parallel in one pass
NN may adopt to the particular training set and lose prediction capabilities

Solution: regularisation techniques

- L2 - regularisation
  \[ C(w, b) \to C(w, b) + \lambda \left( \sum w^2 + \sum b^2 \right) \]
- admixture of random noise to data
- Dropout (see appendix)
Dealing with sequences/images

MLP are not very effective when dealing with high-dimensional data

+ Large input data size
+ Signals in the adjacent pixels are often correlated
+ We want more layers - implement more complex logic
+ We want less weights - easier to train

Convolutional layers

• Small amount of shared nonzero weights (kernel)

\[ \sigma \left( b + \sum_{l=0}^{4} \sum_{m=0}^{4} w_{l,m} a_{j+l,k+m} \right) \]

• Capture local feature maps
Pooling layer

Scale down the image to capture larger-scale features

Operation:
- Maximum
- Average
- ...
Convolutional NN architecture

Minimal

Galaxy zoo challenge winner

morphological classification of galaxies based on images