Muon RNN with TA SD

~~Apr. 25, 2024~~ → Oct. 15, 2025

Keitaro Fujita(ICRR)

Motivation

- Advantage of muon measurement by SD array is:
    - mass composition study with huge statistics even if highest energies
    - invisible energy estimation
    - muon puzzle
        - energy scale ??

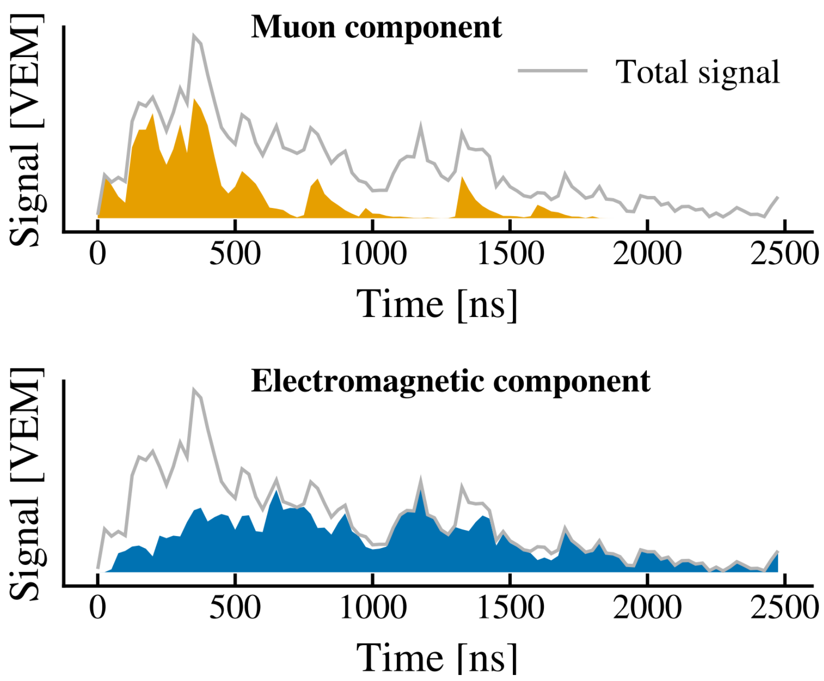
- TA SD is sensitive for both muon and electromagnetic component
    - this is advantage and disadvantage

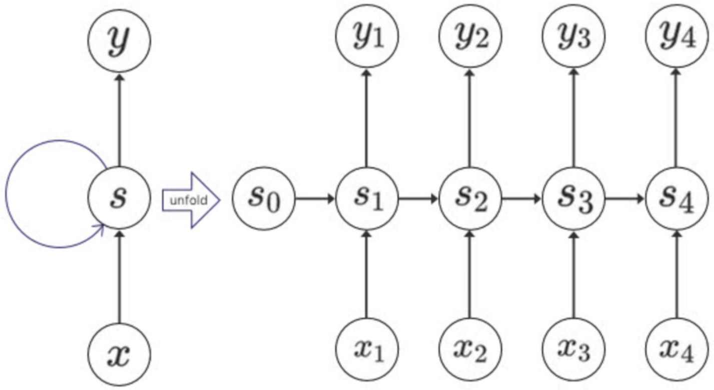
- But our SD waveforms should have info. for muons
    - Early part
    - (sharp shape)
    - Coincidence with both layers

Introduction of Auger study

- Auger SD
  - 3 PMTs are installed in a SD
  - PMT output is digitized by 40MHz (25ns bins), 10bit FADC
    - signal trace has a length of 768 samples(=19.2μs)
  - Muon separation
    - inclined showers, distances far from shower core

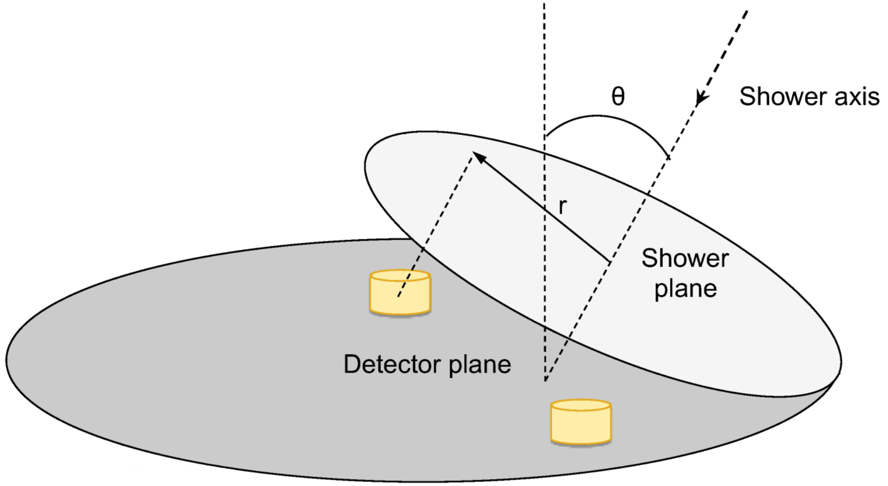
- Machine learning approach for separation $\boldsymbol{\mu}$/em.
  - Recurrent Neural Networks (RNNs) is used





Auger study: RNNs

- Inputs
  - calibrated first $5\mu s$ signal trace in VEM unit
    - 3 PMT signals are averaged
    - $E < 10^{19}$ eV
      - 90% SDs has complete μ signal in $5\mu s$
        Remaining SDs has 99% μ signal in $5\mu s$
    - $E > 10^{19}$ eV
      - 70% SDs has complete μ signal in $5\mu s$
        Remaining SDs has 99% μ signal in $5\mu s$
  - event zenith angle, $\boldsymbol{\theta}$
  - distance far from shower axis, $\boldsymbol{r}$
  - ~~shower energy~~, do not help to improve
    - need to match the energy scale of data if use



RNN architecture

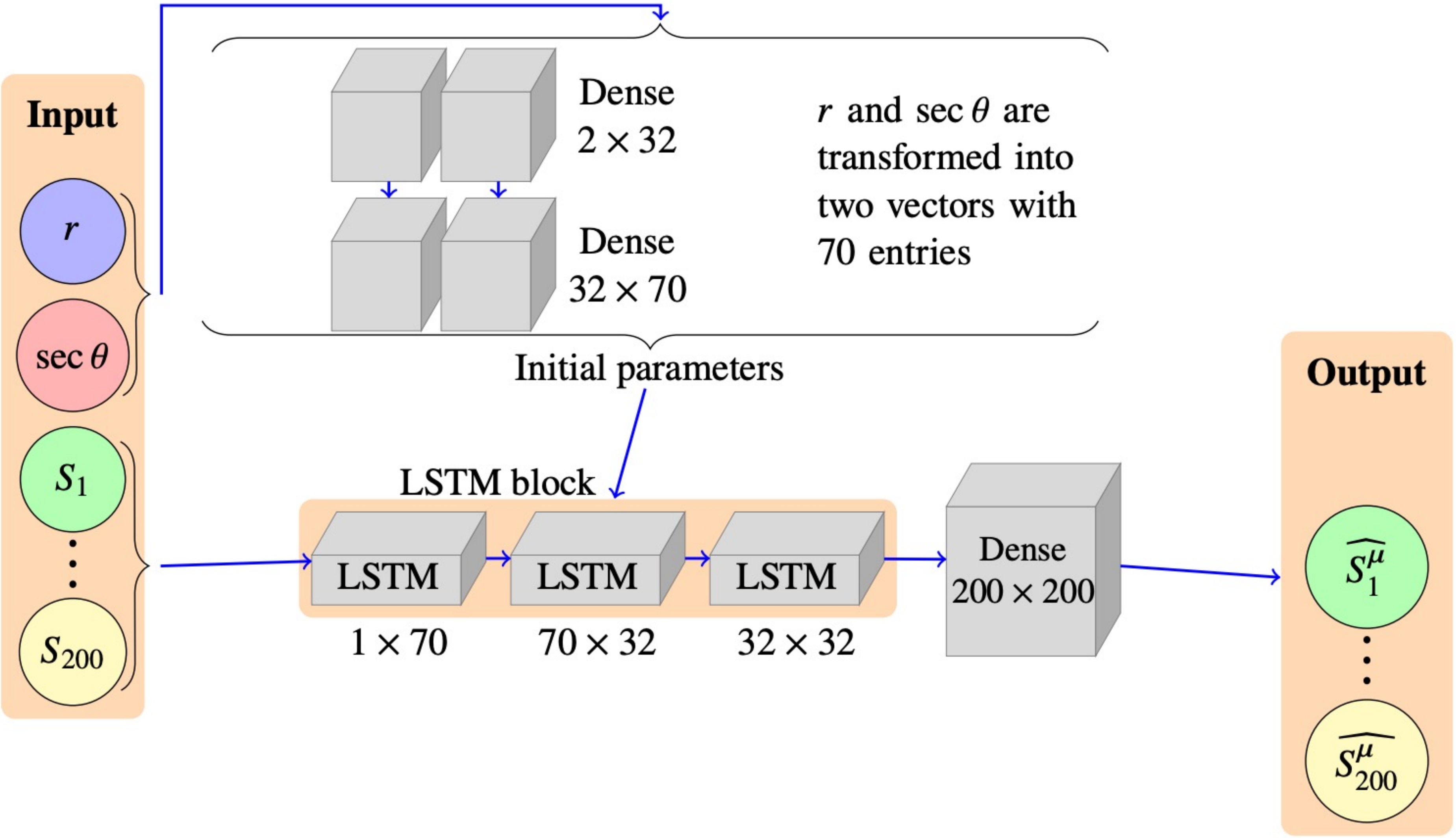
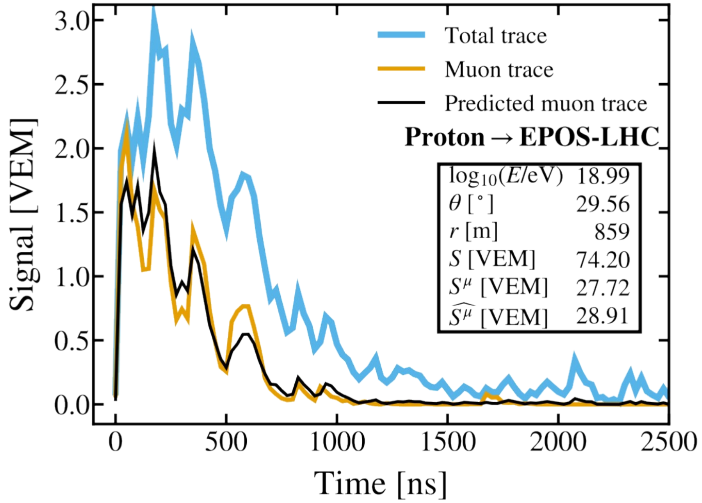
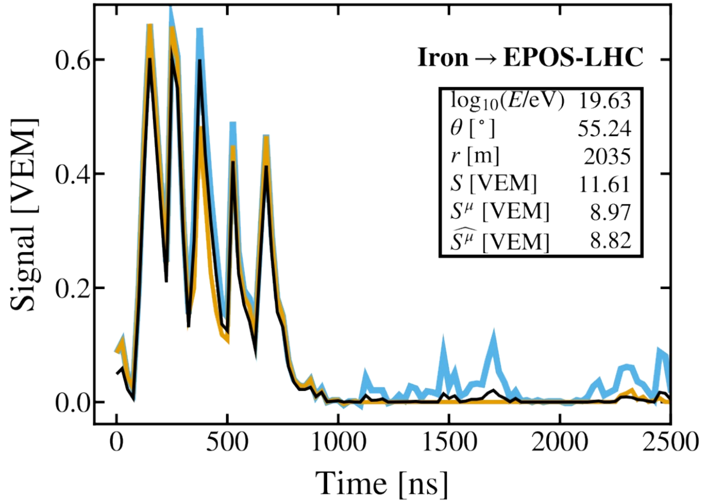


Figure 3. Schematic drawing of the input, architecture and output of the neural network. See the text for details.

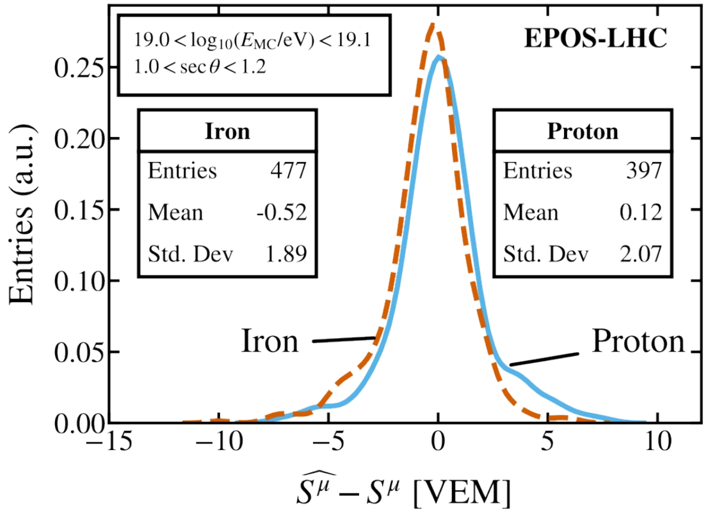
Results

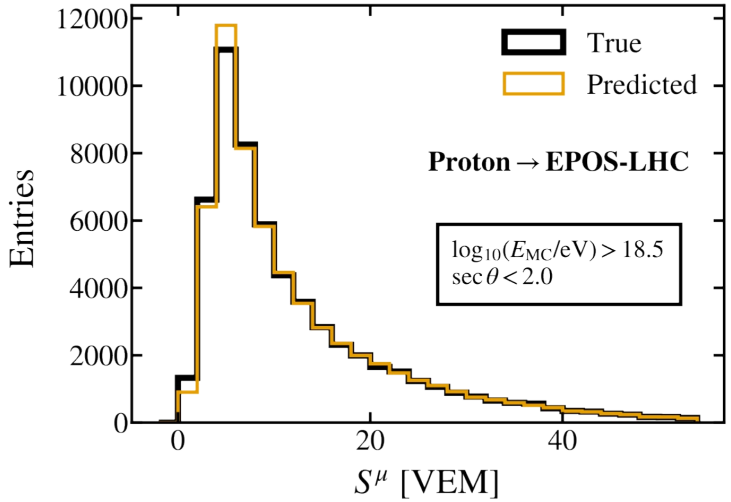
- Predicted muon traces examples





- $\Delta S^{\mu}, \widehat{S_i^{\mu}}$ and S_i^{μ} distribution





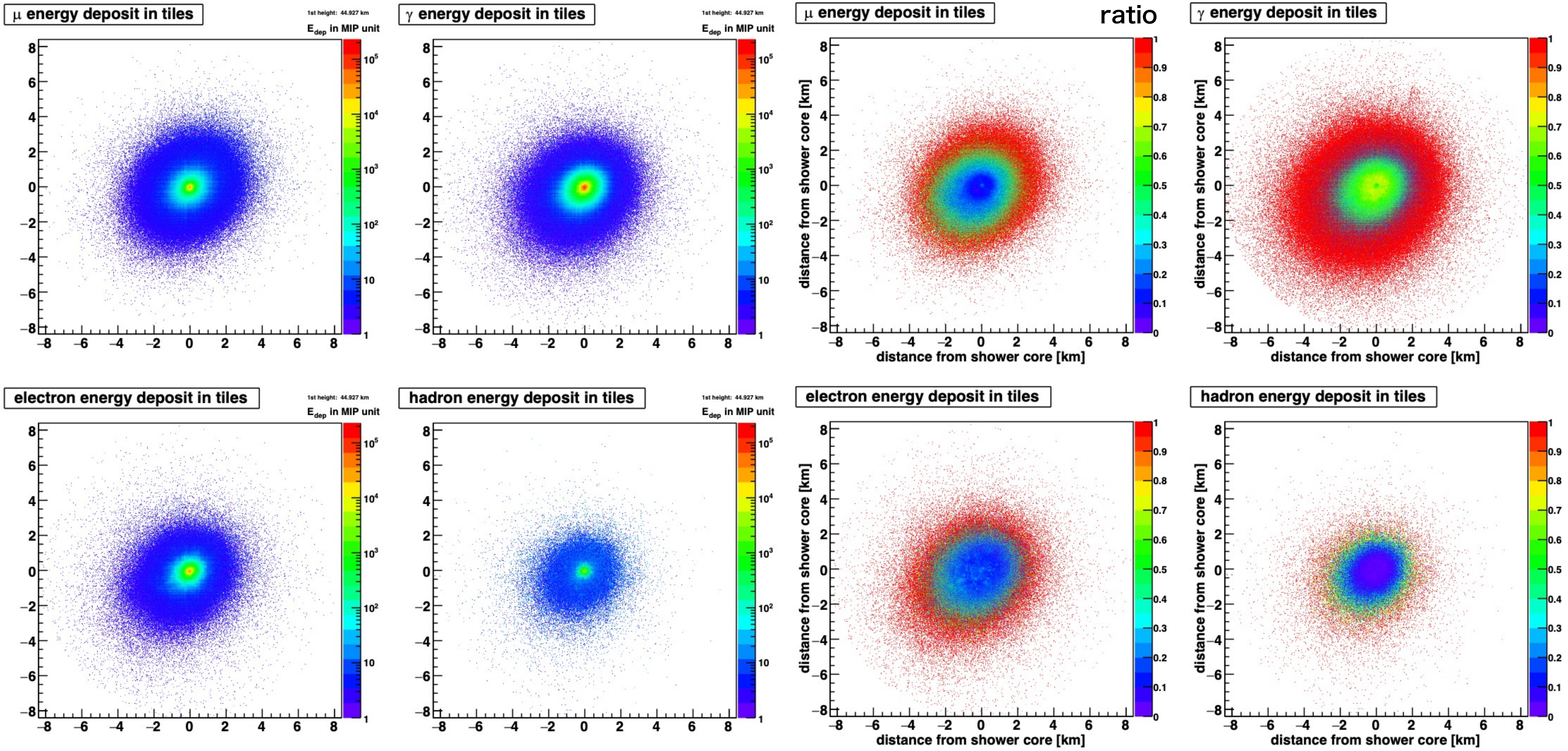
Muon extraction with TA SD

- **We should calculate muon signal in our detector simulation**
    - basically, this had done by Takeishi-san in his previous study
        - QGSJETII-03 was used
    - but as least for me I don’t have his analysis tools
    - If don't have it, let make it

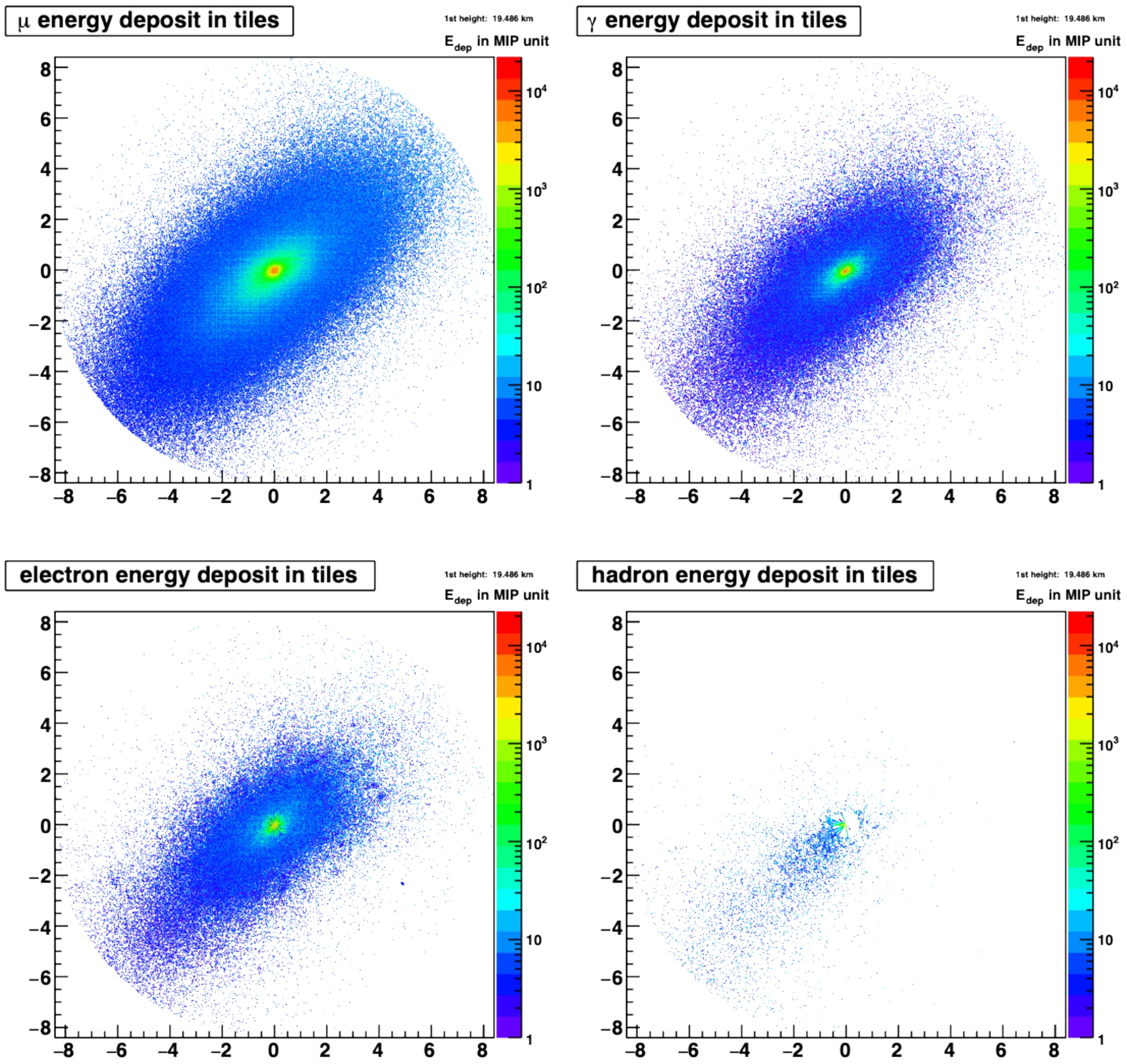
Tile file modification

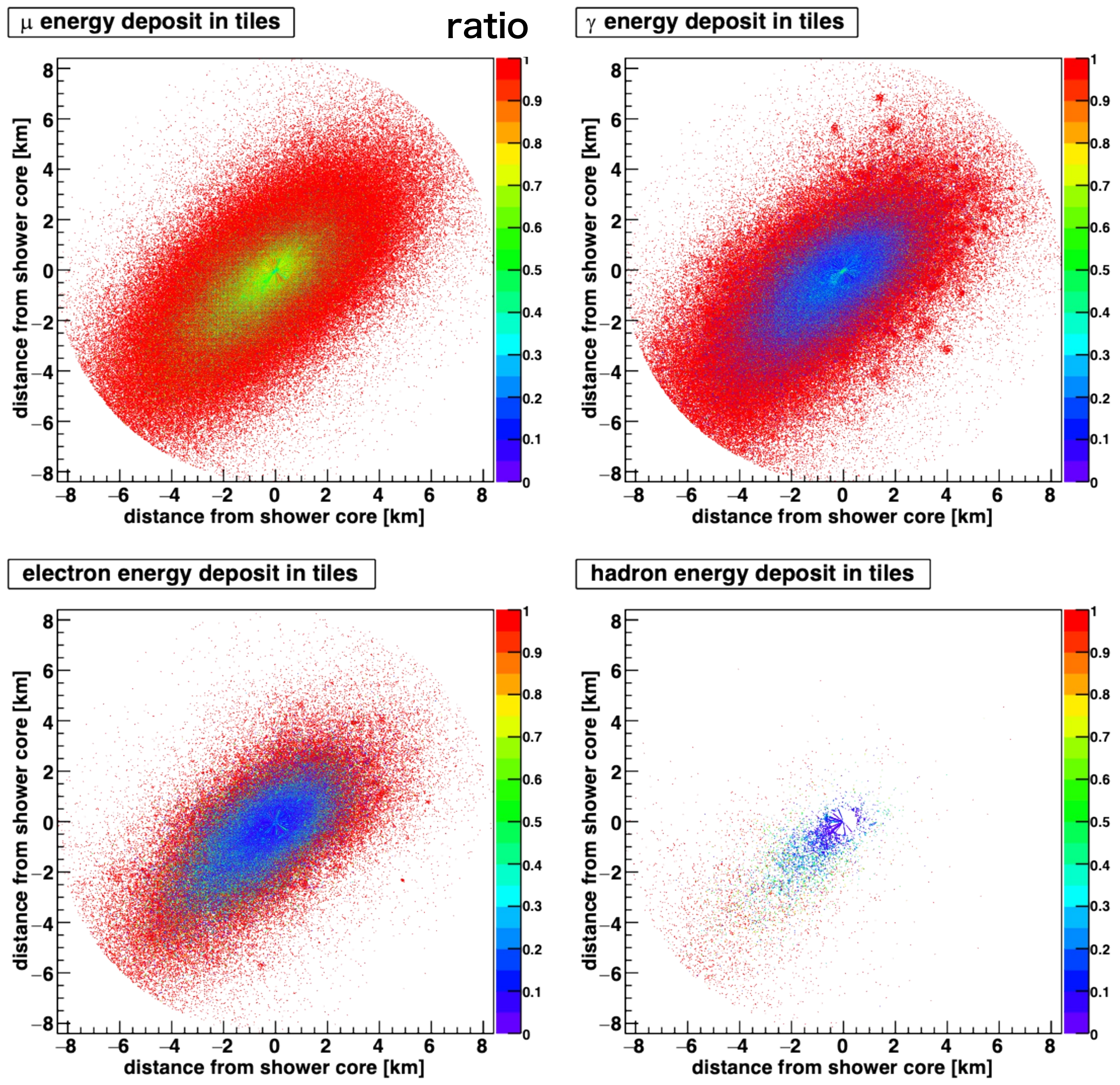
- Data brock structure of original tile file is:
  - event header sub-block: float × 273
    - same as CORSIKA one
    - particle id, total energy, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ ...
  - energy deposition sub-block**: short** × **6**
    - **tile position**, X and Y
    - **upper/lower energy deposition**
    - time
    - momentum in z?
- Add energy deposition by $\boldsymbol{\mu}^{\pm}/\boldsymbol{e}^{\pm}/\boldsymbol{\gamma}$**/others to original tile file**
  - **additional part:** short × **8**
    - **upper/lower energy deposition by** $\boldsymbol{\mu}^{\pm}$
    - **upper/lower energy deposition by** $\boldsymbol{\gamma}$
    - **upper/lower energy deposition by** $\boldsymbol{e}^{\pm}$
    - **upper/lower energy deposition by others**
- File size: ~ x5

Tile file modification ($E = 10\mathrm{EeV}, \theta = 35.3°$)



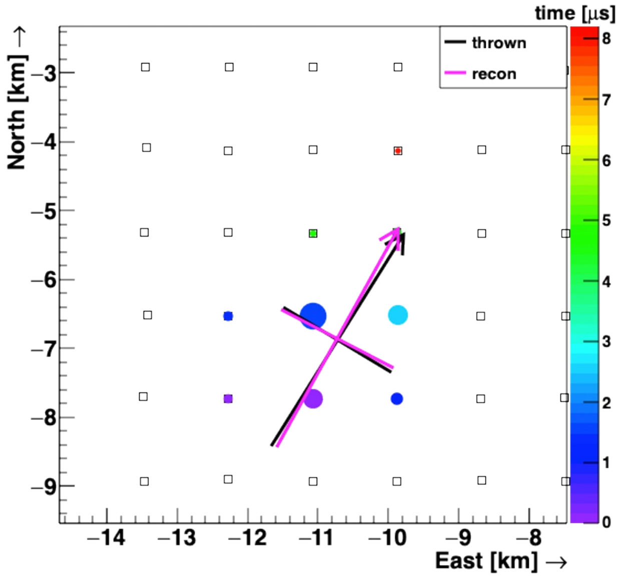
Tile file modification ($E = 10\mathrm{EeV}, \theta = 66.8°$)

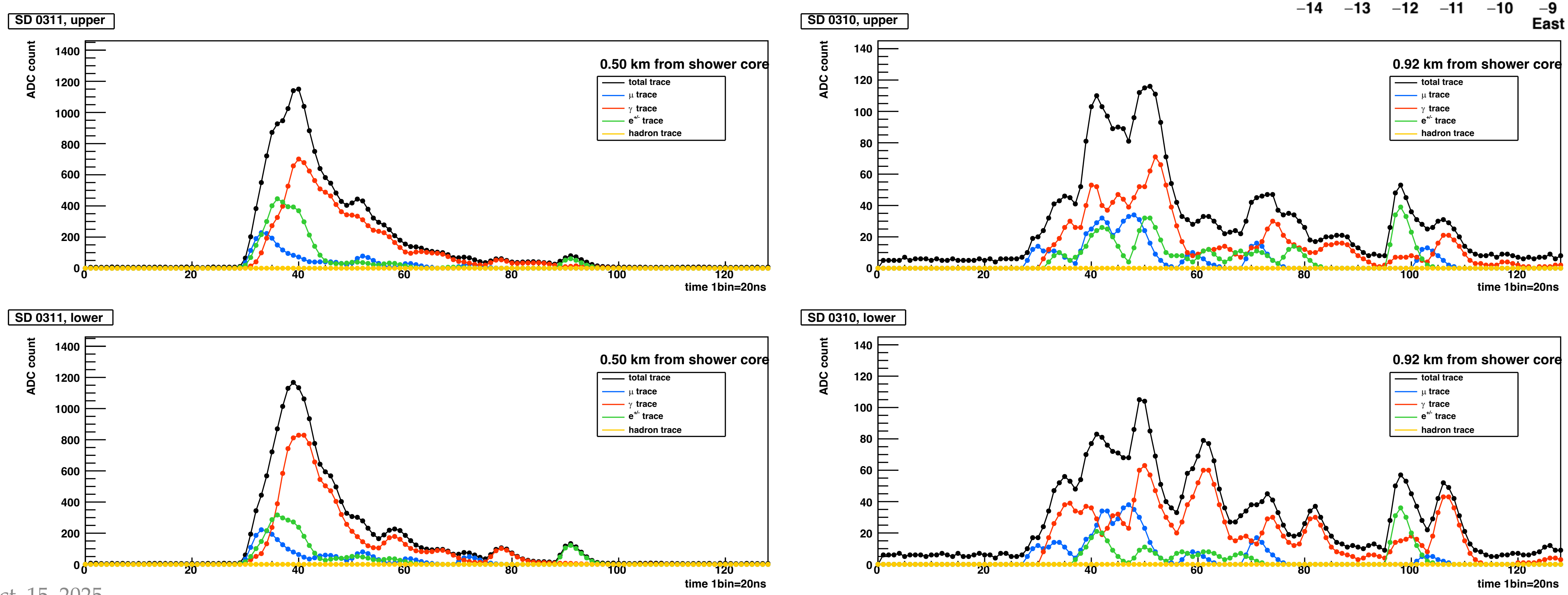




μ-trace of TASD MC

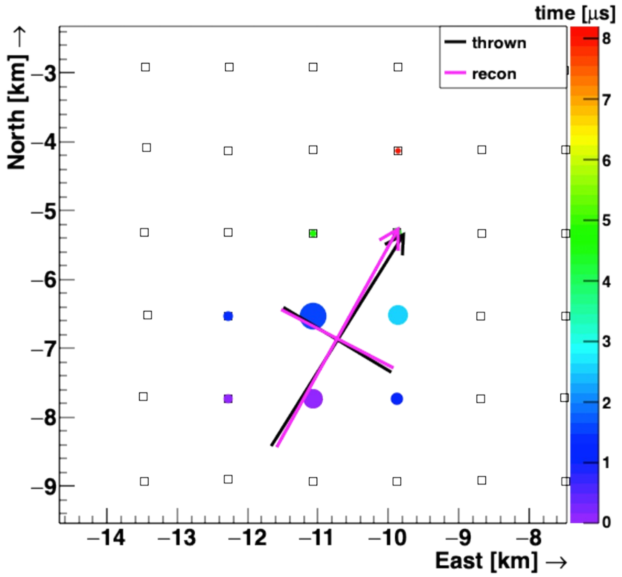
- Calculate waveforms by $\mu^{\pm}/e^{\pm}/\gamma/$others
  - Stored in new four dstbanks
    - default: rusdmc(shower info.), rusdraw(SD waveform)
    - addition: same format of rusdraw but for $\mu^{\pm}, \gamma, e^{\pm}$, and others
- Visualize individual components trace

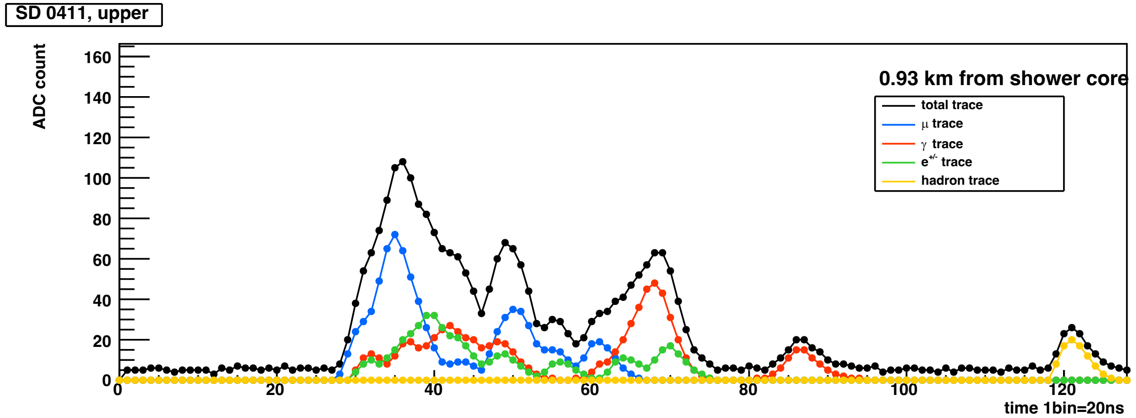


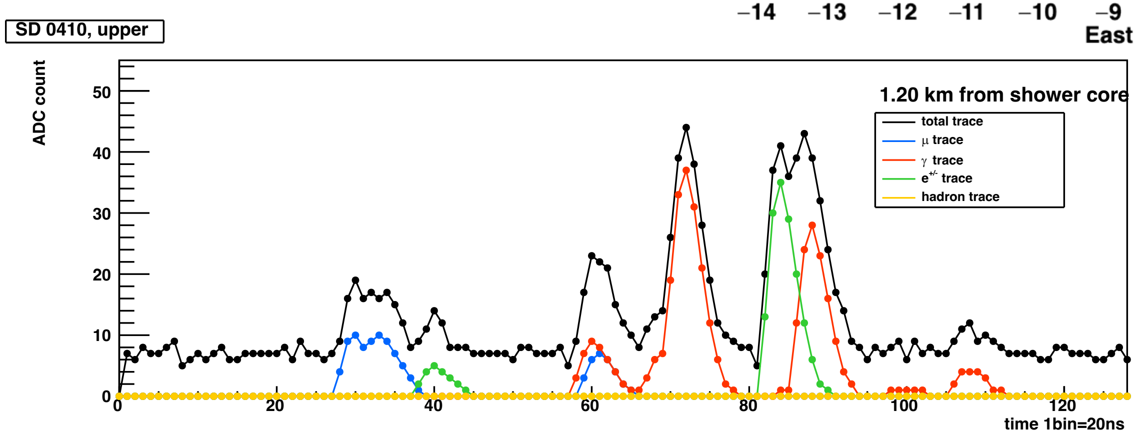


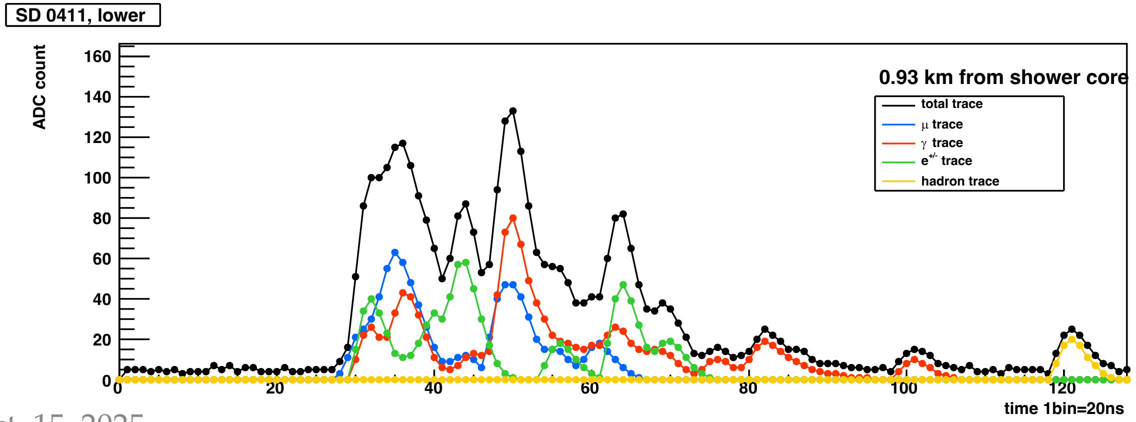
μ-trace of TASD MC

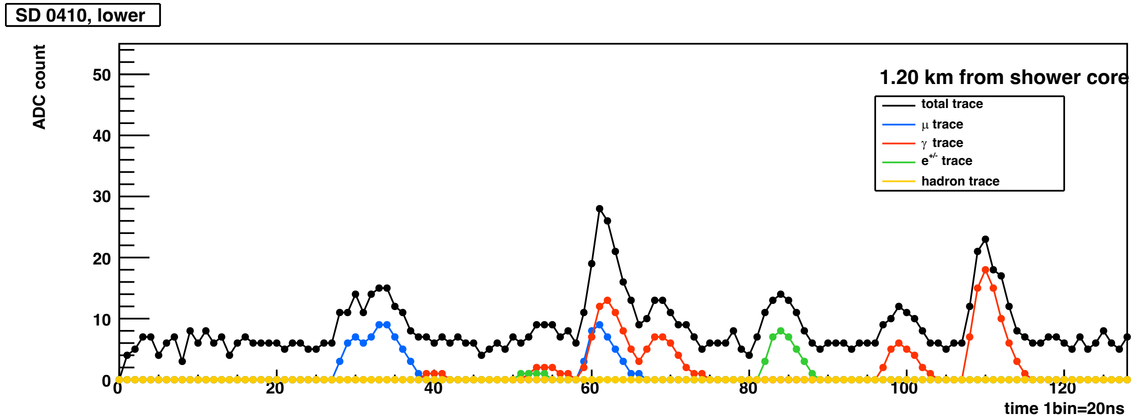
- Calculate waveforms by $\mu^{\pm}/e^{\pm}/\gamma/$others
  - Stored in new four dstbanks
    - default: rusdmc(shower info.), rusdraw(SD waveform)
    - addition: same format of rusdraw but for $\mu^{\pm}, \gamma, e^{\pm}$, and others
- Visualize individual components trace











μ MC dataset, inputs for RNN

- CORSIKA conditions
  - interaction model: QGSJetII-04
  - primary: proton and iron
  - energy range: 10^{19} eV, 10^{20} eV (mono-energtic showers)
  - zenith angle: 0 - 70 deg
- detector simulation with TA SD array (14yrs of calibration data used)
- TA SD standard reconstruction had performed (reconstructed by sdanalysis chain)
  - Events should be passed TASD spectrum cuts except for zenith angle cut
    - $\boldsymbol{\theta}_{\mathbf{rec}}$ < 60deg applied this work (spectrum cuts: $\boldsymbol{\theta}_{\mathbf{rec}}$ < 45deg)
  - Signal selection for inputs
    - Saturated signals are removed
    - signal size > 1.5 VEM
  - Only first frame of waveform is used
  - Total ~600000 waveforms for training
- Inputs for network
  - sec$\boldsymbol{\theta}$, distance from core
  - 128 bins waveform, both layers

Data process

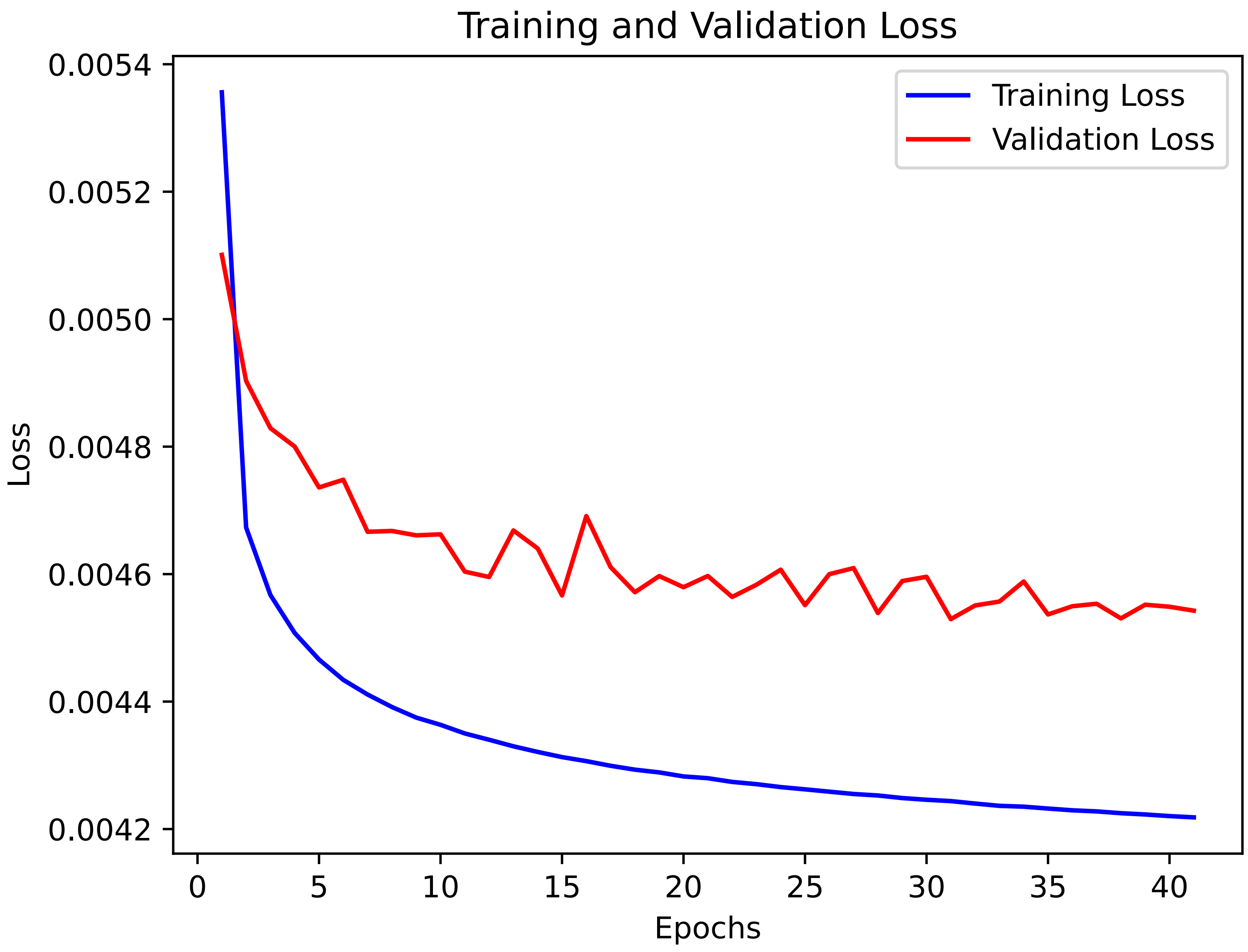
- RNN architecture: mimicked Auger work
- Distance $\boldsymbol{r}$ and $\mathbf{sec}\boldsymbol{\theta}$ are scaled to be between 0 and 1
- All traces are scaled individually to be between 0 and 1
- True (simulated) muon trace is scaled with the same factor used for the total trace
  - Both **upper/lower trace using (NOT summation), 256 datapoints in total**
    - give two dimensions array
- Output of the neural network is also between 0 and 1 so as to use the same factor to rescale back the predicted trace

$$\boldsymbol{L} = \frac{\mathbf{1}}{\mathbf{128\times 2}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{128\times 2}} \left(\widehat{\boldsymbol{S}_{\boldsymbol{i}}^{\boldsymbol{\mu}}} - \boldsymbol{S}_{\boldsymbol{i}}^{\boldsymbol{\mu}}\right)^{\mathbf{2}}$$

- Loss function:
- Optimizer: ADAM
- Nvidia 3070 Ti GPU, training takes around 0.5-1 hours

Results

- Training and validation loss vs. epochs

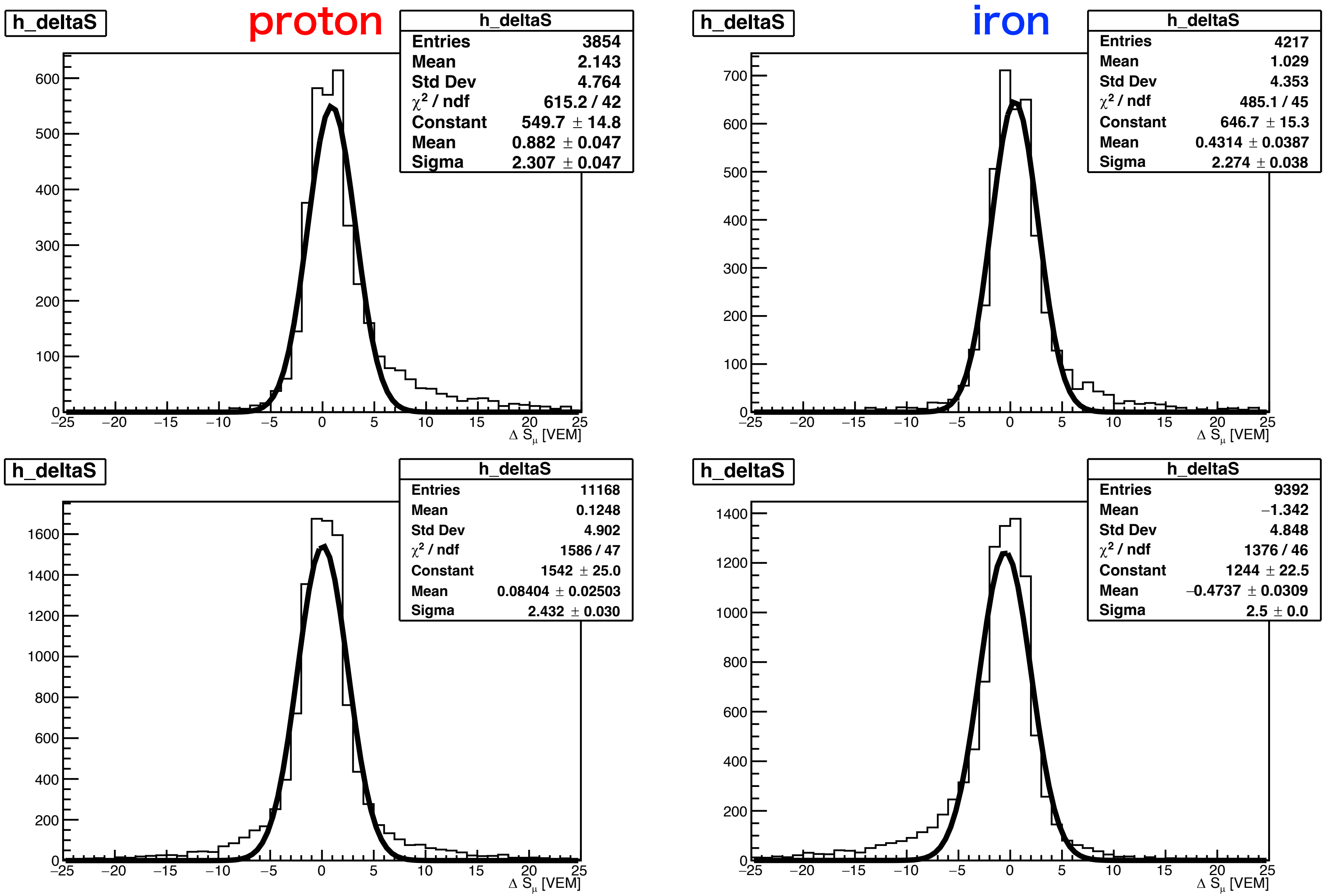


Results

- muon prediction resolution: $\boldsymbol{\Delta S_\mu = S_{\mu,\mathrm{pre}} - S_{\mu,\mathrm{true}}}$ [VEM]

$E = 10^{19}\,\mathrm{eV}$

$E = 10^{20}\,\mathrm{eV}$

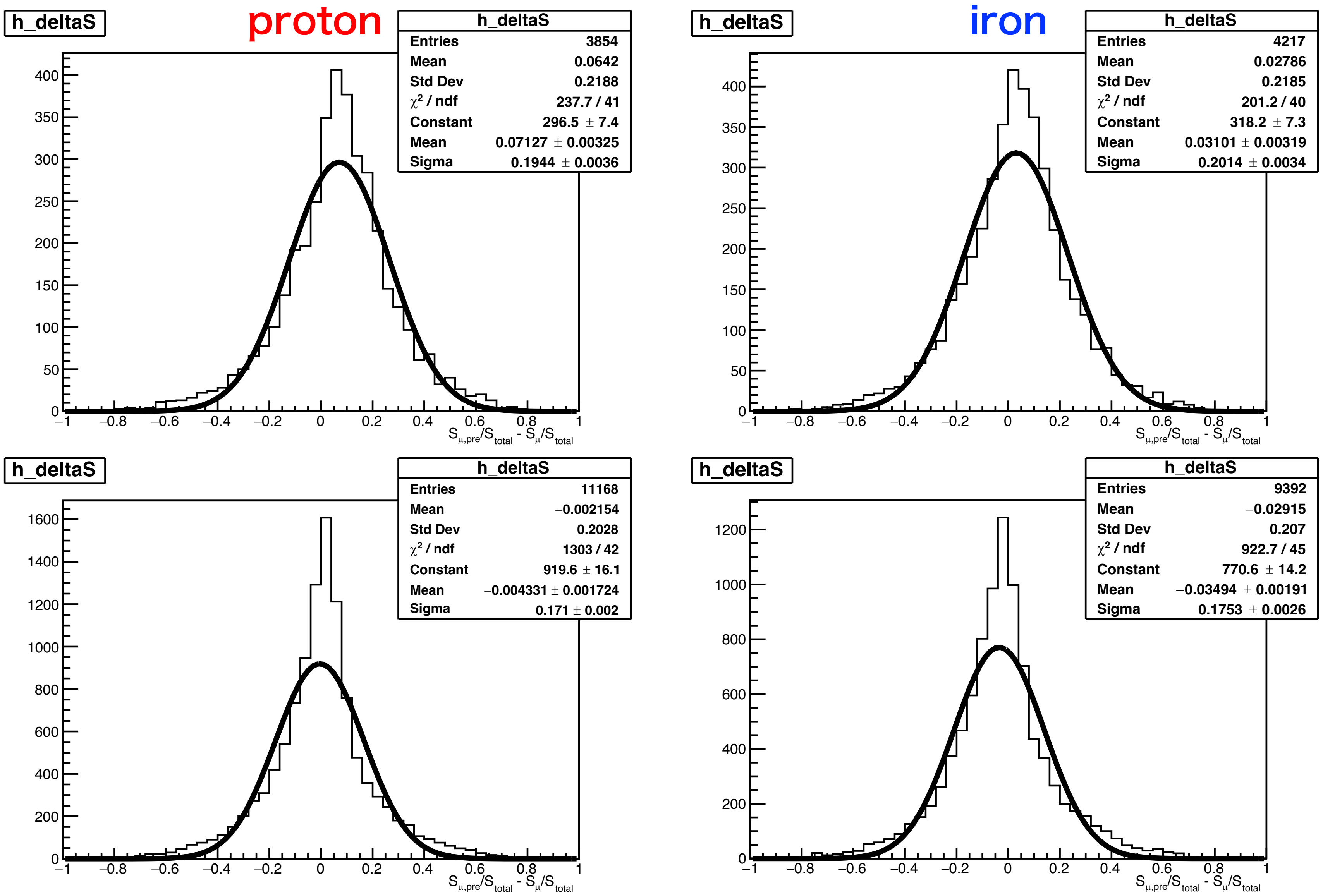


Results

- muon prediction resolution: $S_{\mu,\mathrm{pre}}/S_{\mathrm{total}} - S_{\mu,\mathrm{true}}/S_{\mathrm{total}}$ [%]

$E = 10^{19}\,\mathrm{eV}$

$E = 10^{20}\,\mathrm{eV}$

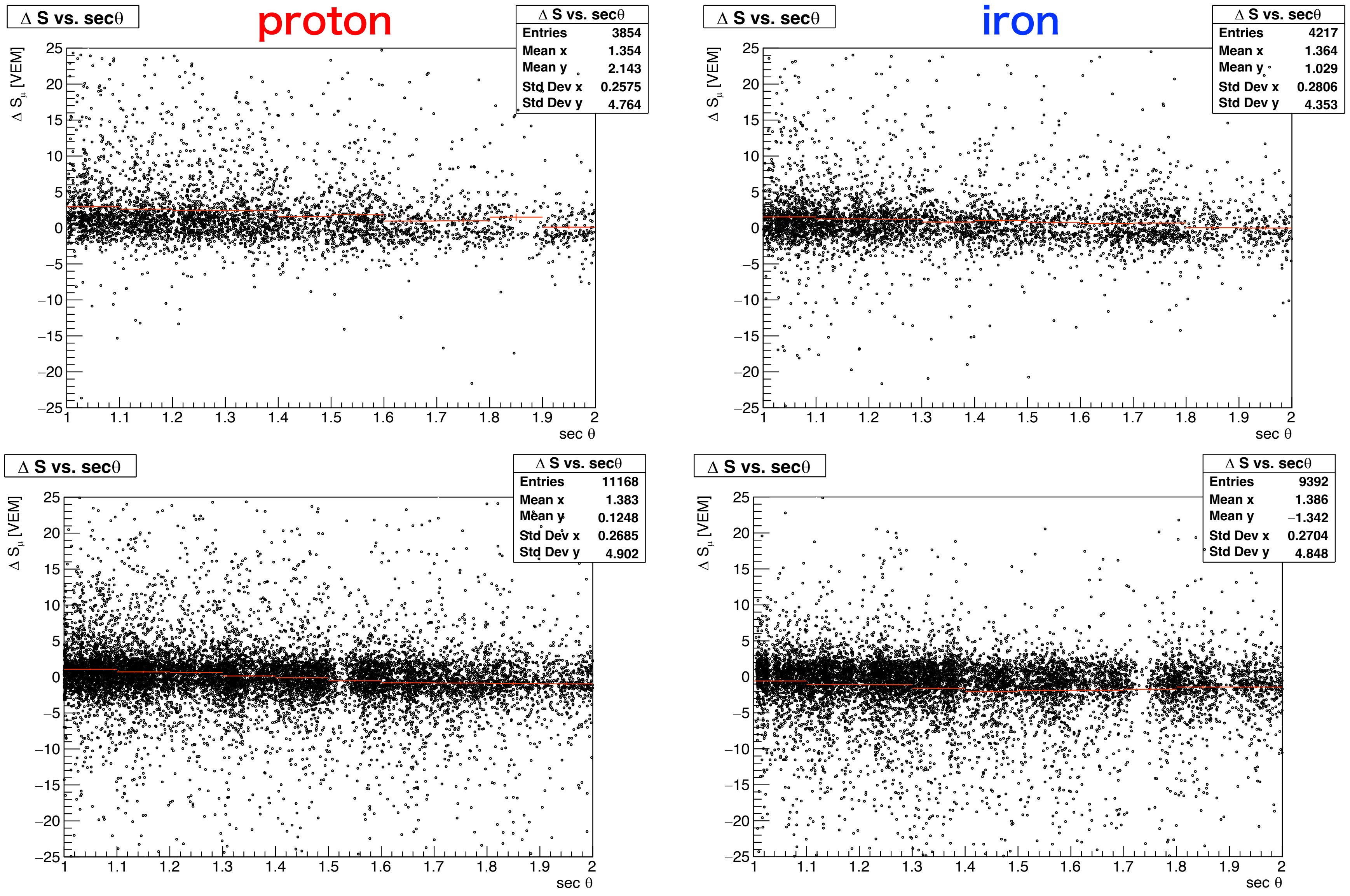


Results

- muon prediction, ΔS_μ vs. $\sec\theta$

$E = 10^{19}$ eV

$E = 10^{20}$ eV



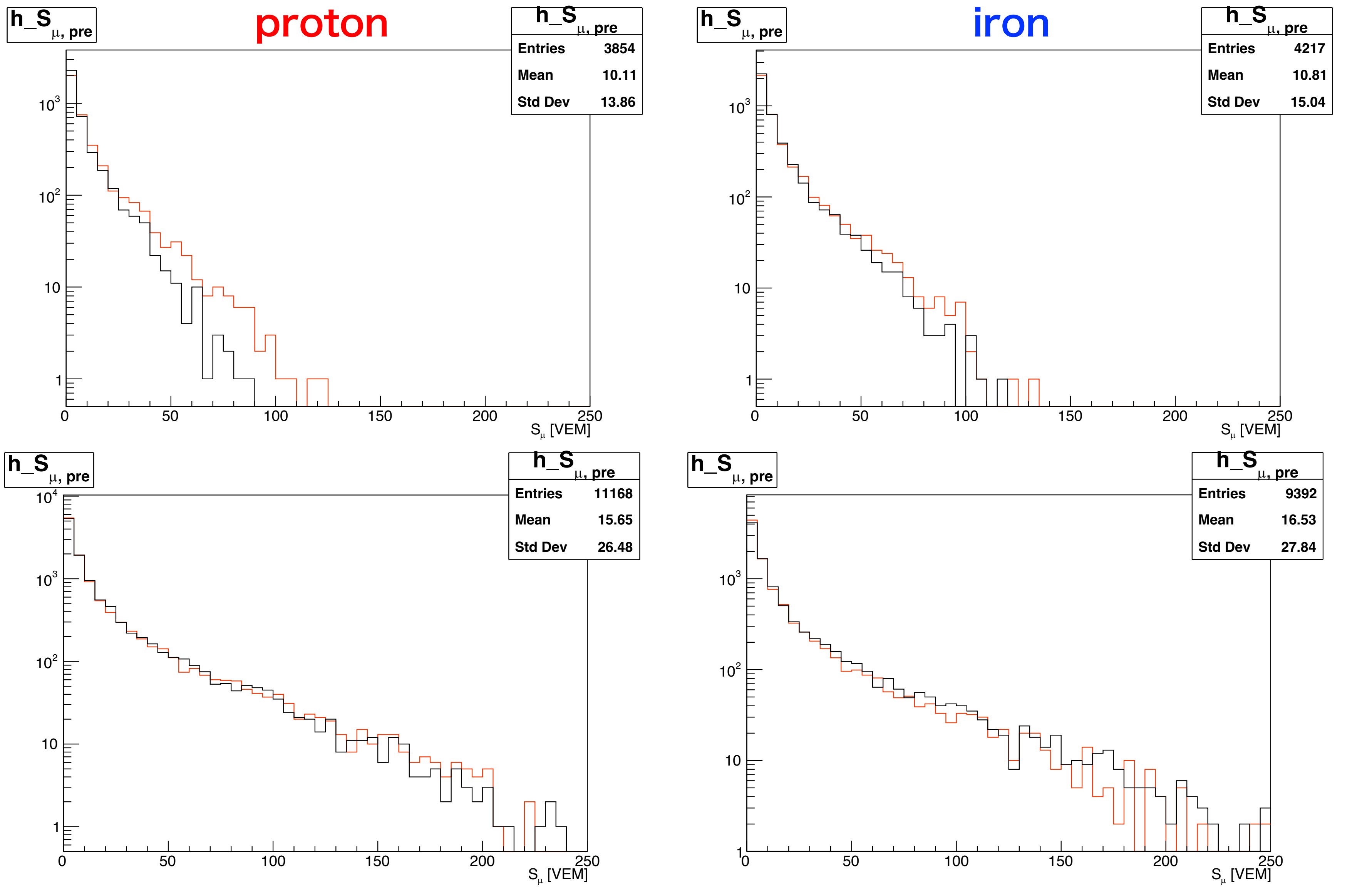
Results

- muon prediction, μ signal distribution

- prediction
- true

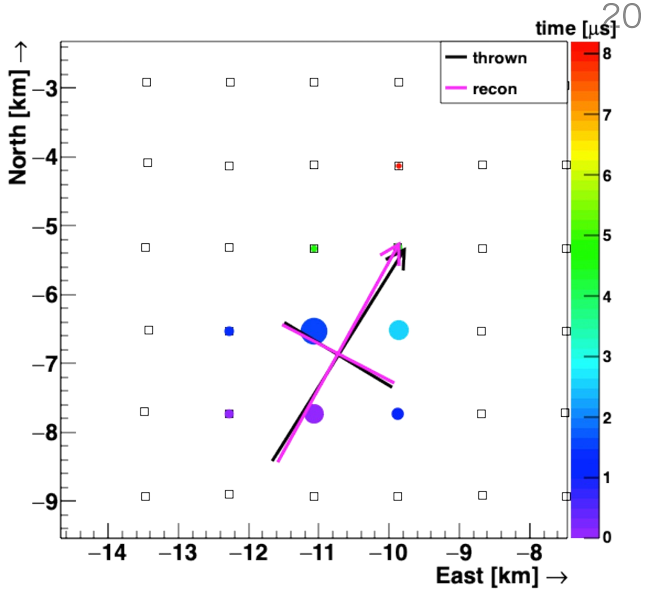
$E = 10^{19}$ eV

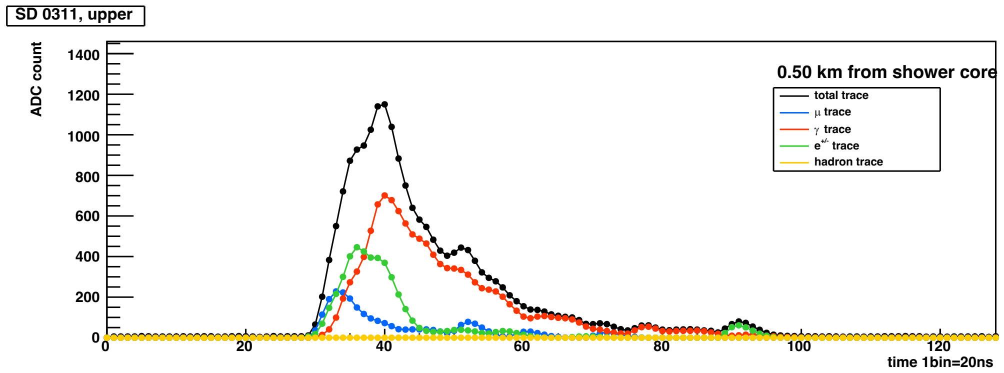
$E = 10^{20}$ eV

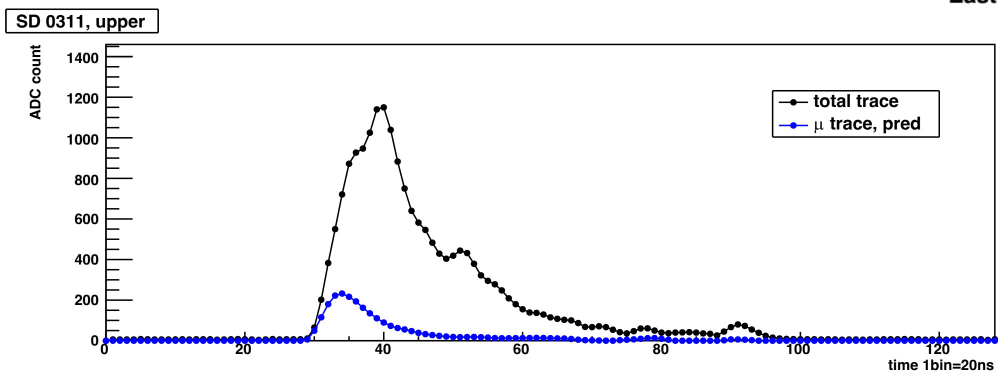


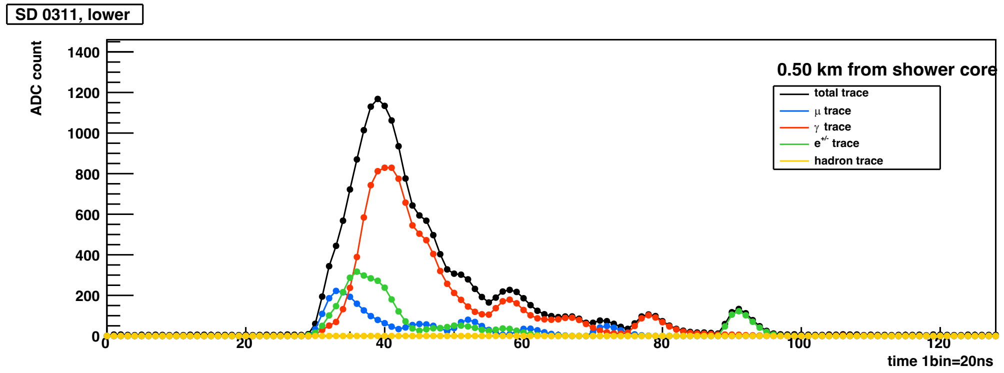
Results

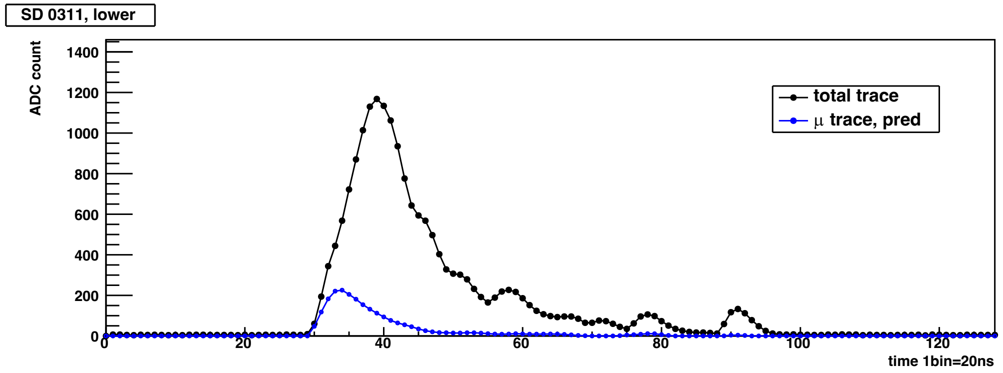
- Muon signals comparisons
  - true (left) vs. prediction (right)





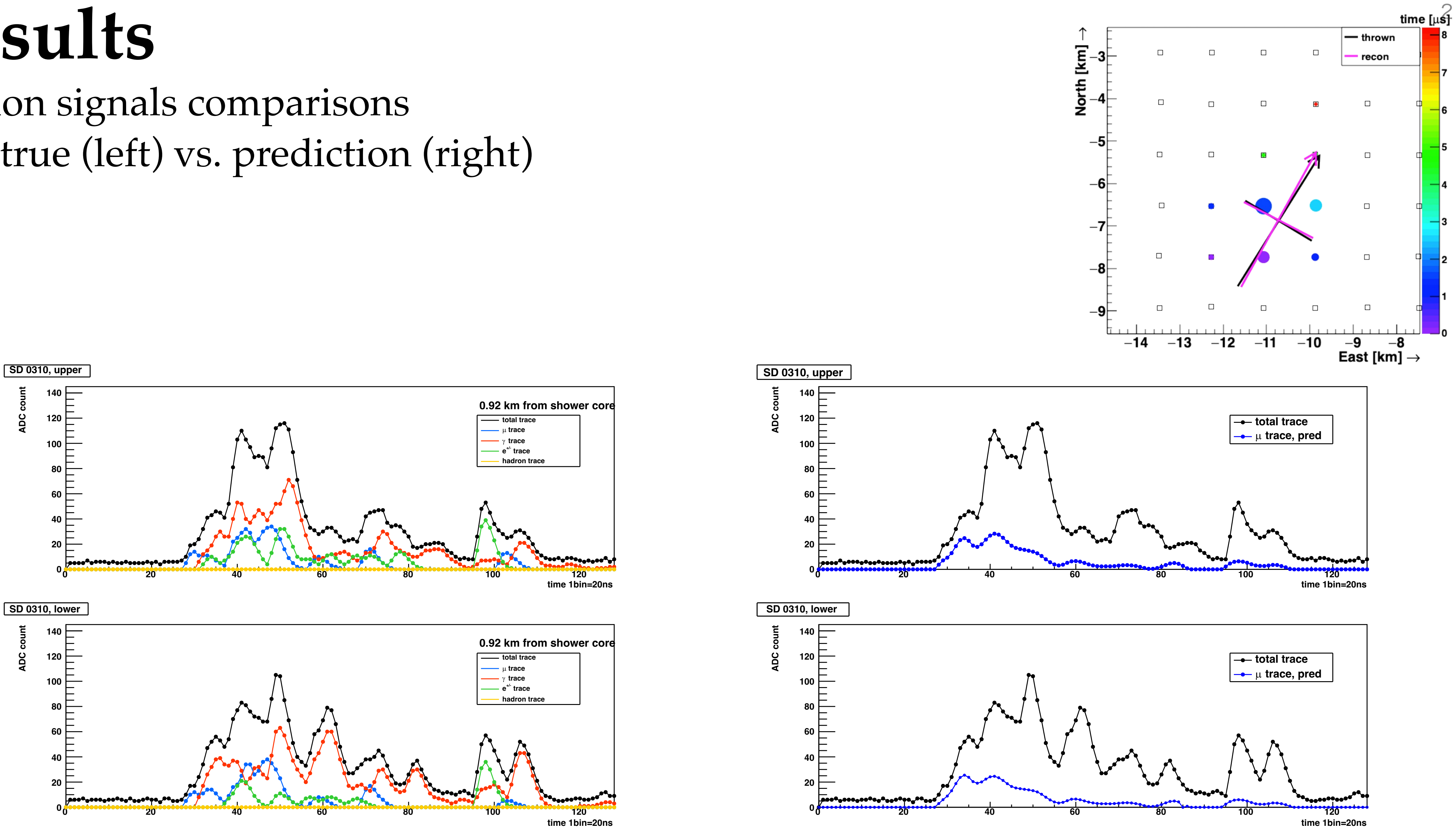






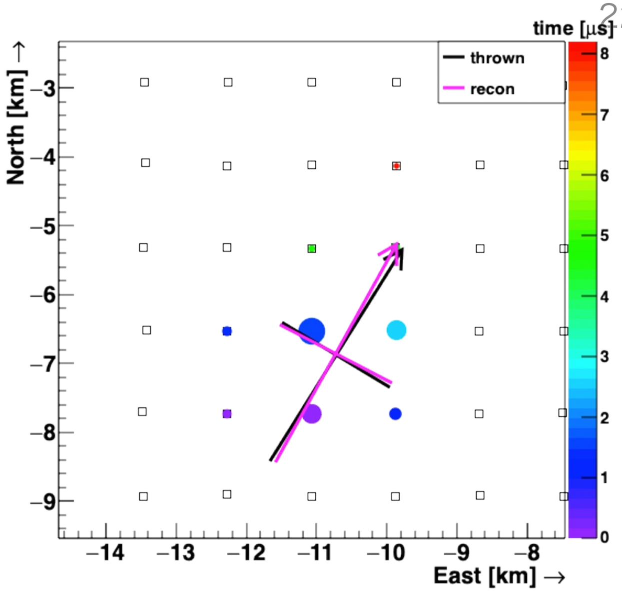
Results

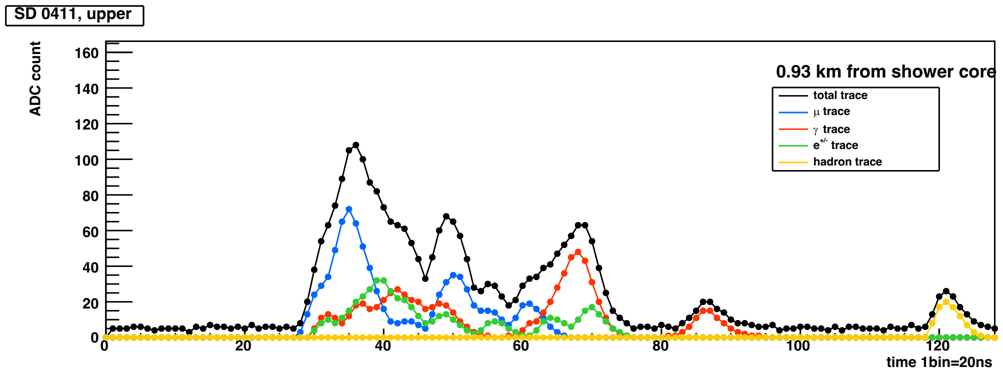
- Muon signals comparisons
    - true (left) vs. prediction (right)

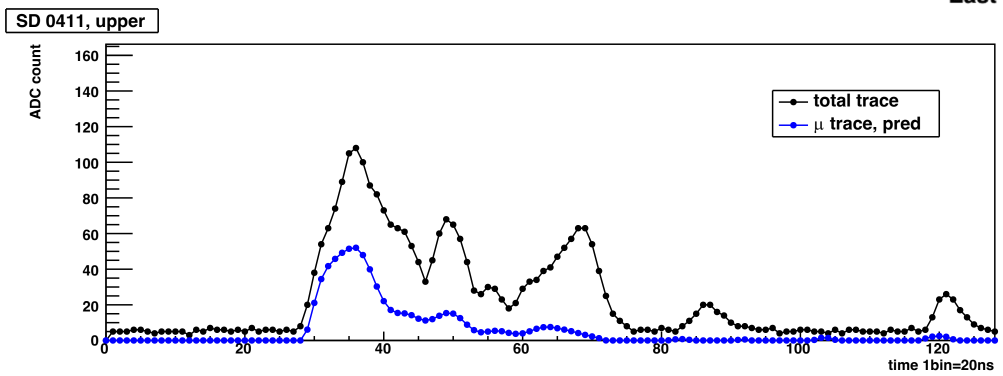


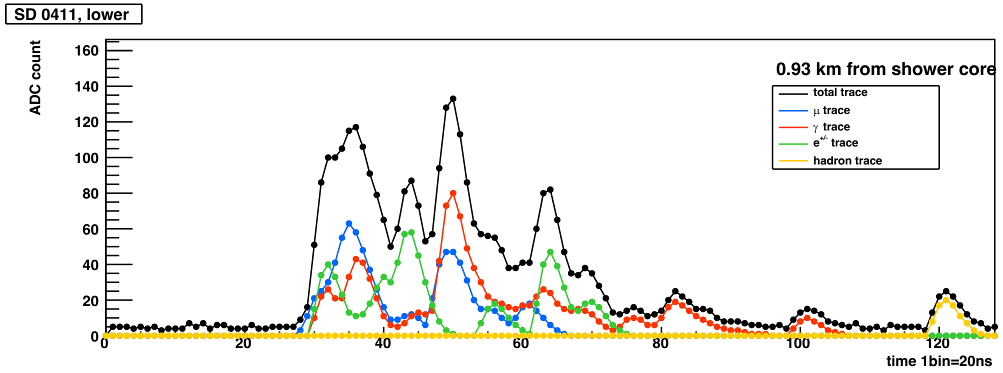
Results

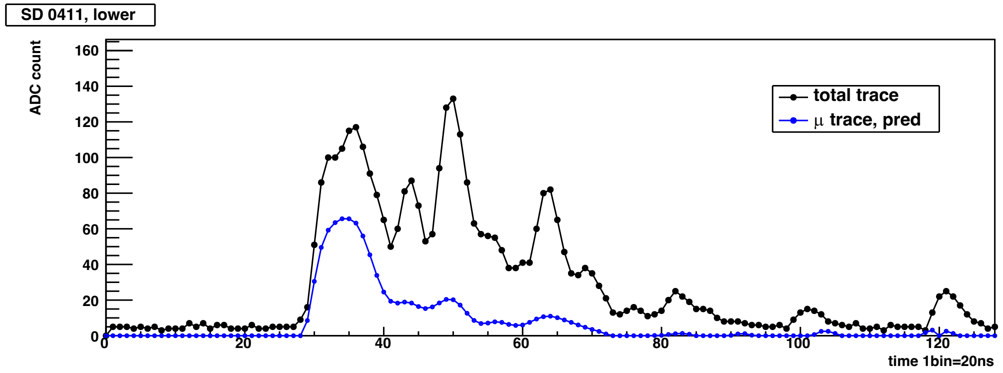
- Muon signals comparisons
  - true (left) vs. prediction (right)





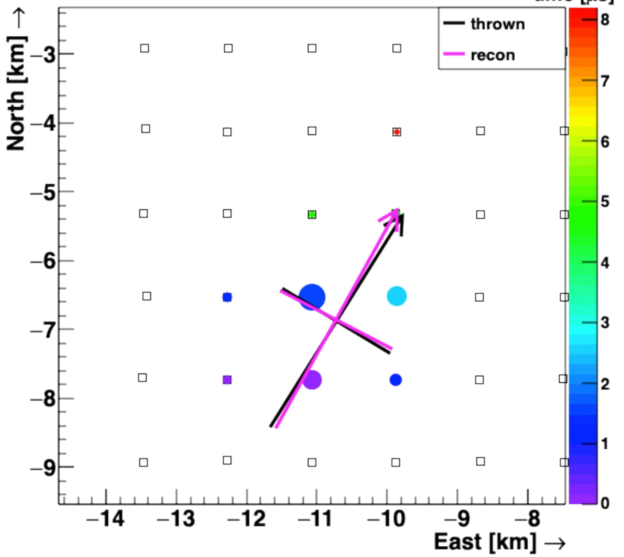


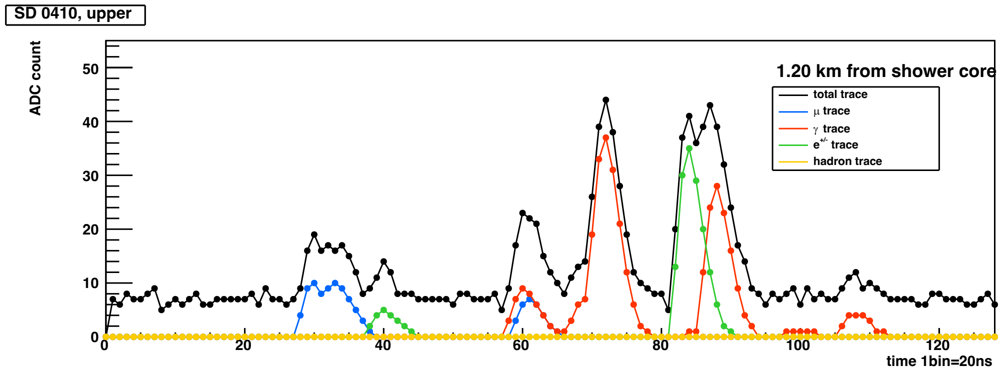


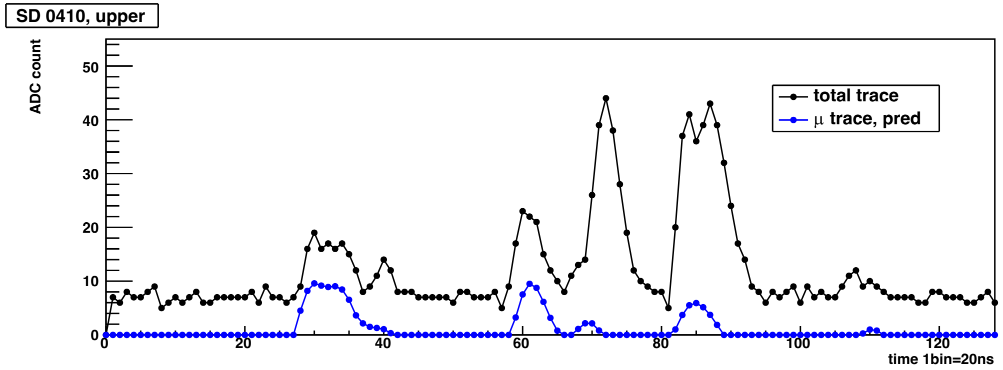


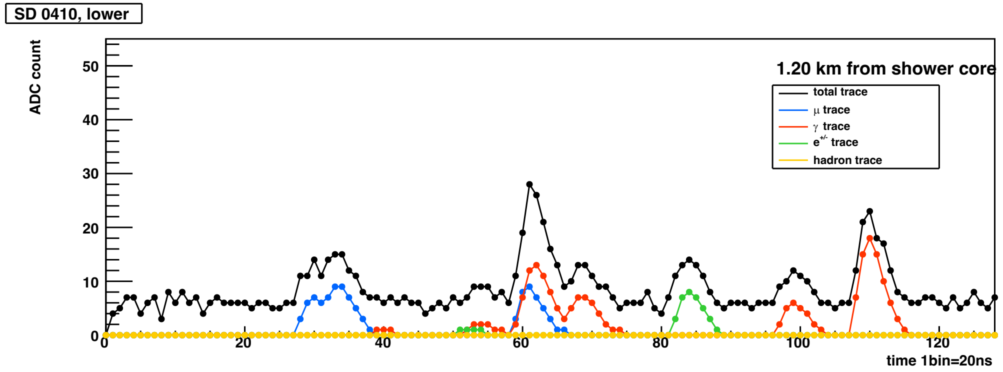
Results

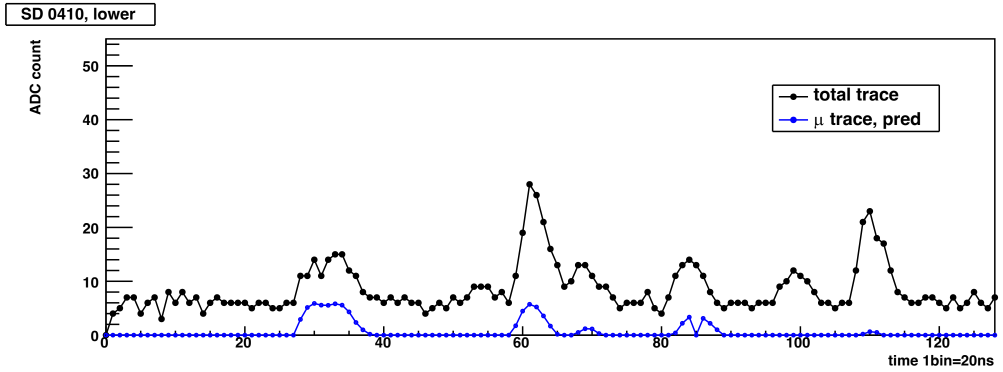
- Muon signals comparisons
  - true (left) vs. prediction (right)









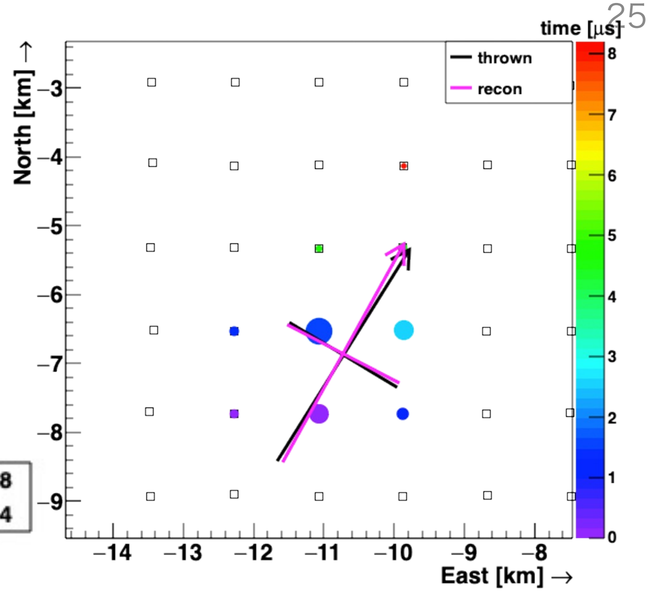


Summary

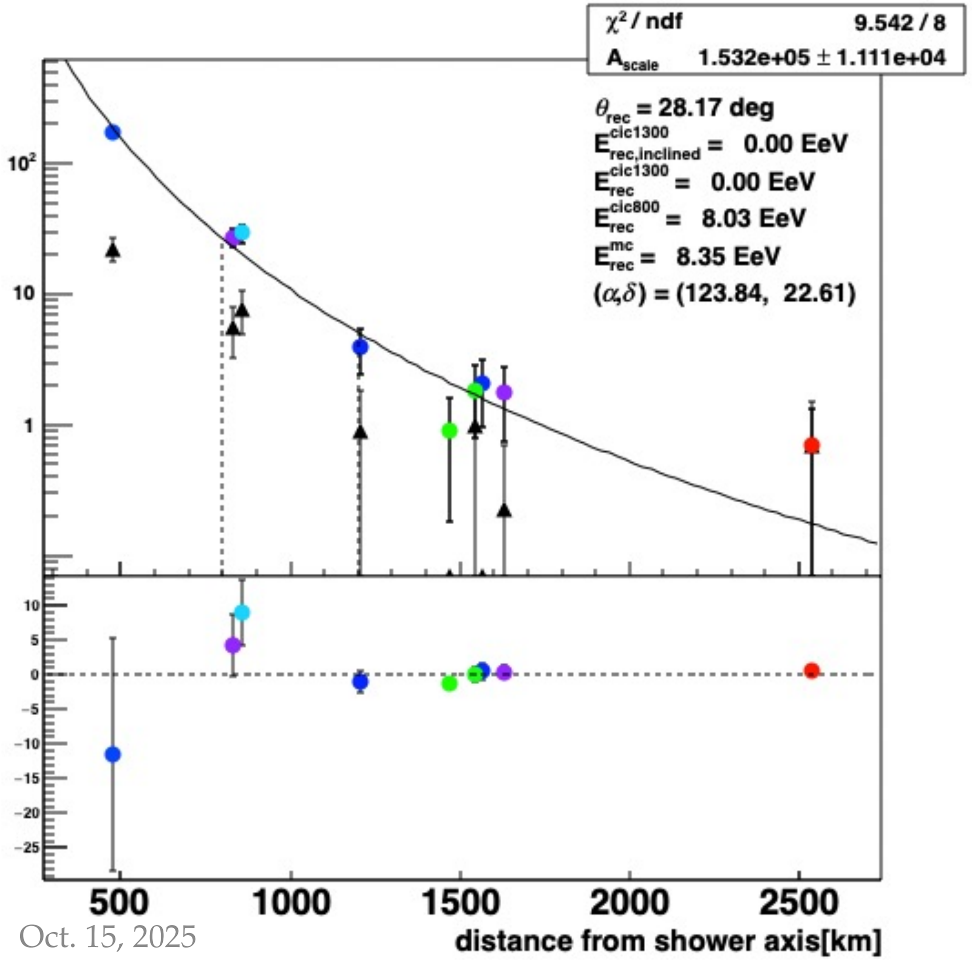
- Muon is important key for
  - Mass composition measurement by SD
  - Invisible energy estimation
  - Muon puzzle
- Create μ-MC with TA SD
  - Modified tile file production, add four-dstbanks
- Shown simulated μ-traces detected by TA SD
- Introduce RNN
- Very preliminary results of extraction of muon signal are shown

Results

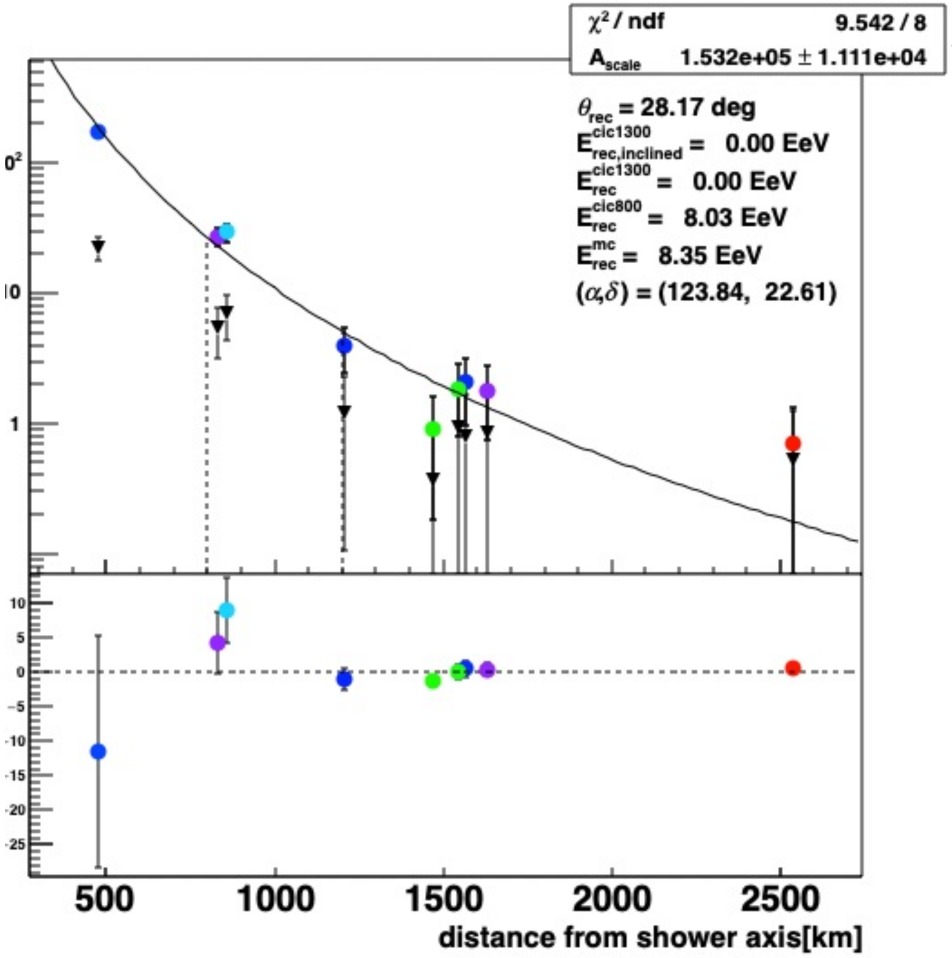
- Muon lateral distribution example



▲: true μ signal



▼: predicted μ signal



it would be nice if we can convert muon lateral dist. to N_μ itself

Comments

Anatoli

- 10:24 AM
- Hi!
- I thought a bit about testing your network. The following steps may help identifying whether the network learns some simple features and proportionalities or if it really can identify something from the signal.
    1. Tran and validate as usual
    2. Create modified test waveforms by multiplying the muon signal by two (or X) and adding it with the remaining signal
    3. Add a sec(theta) or R dependence to the multiplication factor
    4. Remove the muon signal entirely, or make it very little
    5. In case you have access to the EPOS simulation (I remember Sako-san made some a while ago) train on EPOS (train + valid), reconstruct test with network trained QGSJET
- If the network only learns the statistics it won't catch these differences.

Auger study
(for backup)

Data selection and training

- CORSIKA
  - interaction model: EPOS-LHC
  - primary: proton/helium/oxygen/iron
  - energy range: $10^{18.5}$ - $10^{20.2}$ eV
  - zenith angle: 0 - 60 deg

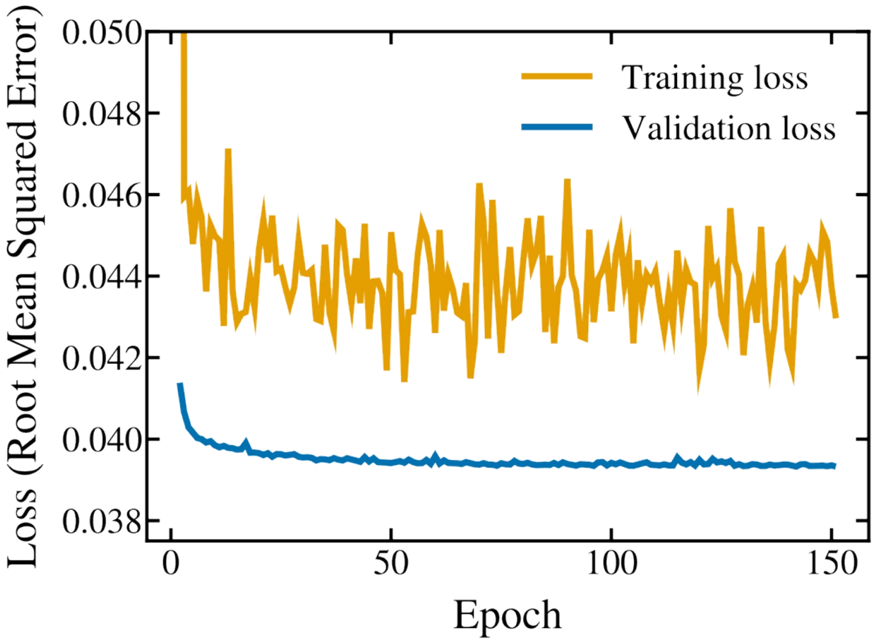
- Only traces for which the integral > 5 VEM are selected
- 450 000 events were available
  - events were sampled randomly and assigned to the training, validation or test data sets using a uniform distribution in energy and secθ for the validation and test sets; the remaining events were assigned to the training set.
  - split as follows:
    390 000 in the training set
    22 000 in the validation set
    34 000 in the test set

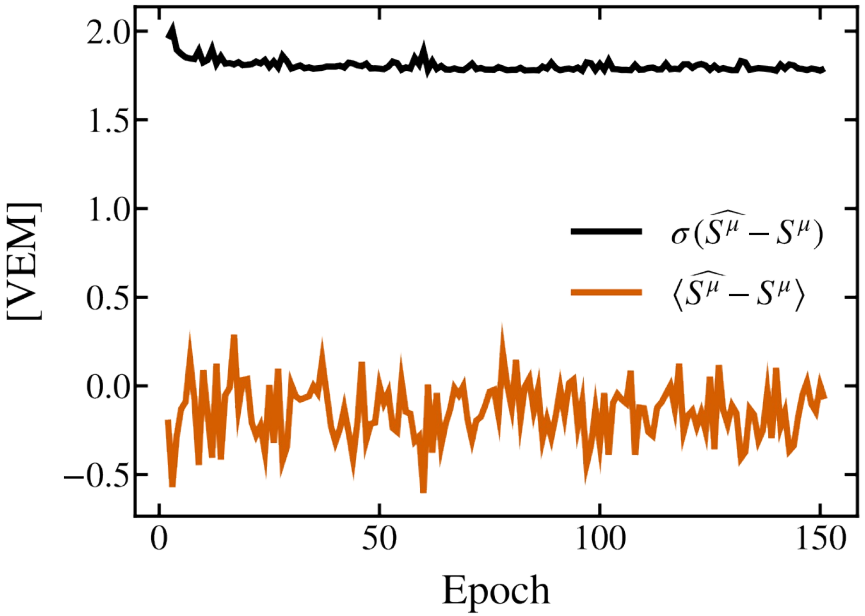
Data selection and training

- distance $\boldsymbol{r}$ **and zenith angle** $\boldsymbol{\theta}$ are scaled to be between 0 and 1
- all traces are scaled individually to be between 0 and 1
- True (simulated) muon trace is scaled with the same factor used for the total trace
- Output of the neural network is also between 0 and 1 so as to use the same factor to rescale back the predicted trace
- loss function: $$\boldsymbol{L = \frac{1}{200}\sum_{i=1}^{200}\left(\widehat{S_i^{\mu}} - S_i^{\mu}\right)^2}$$
- optimizer: ADAM
- Nvidia 2080 Ti GPU, the training takes around 8 hours

Data selection and training

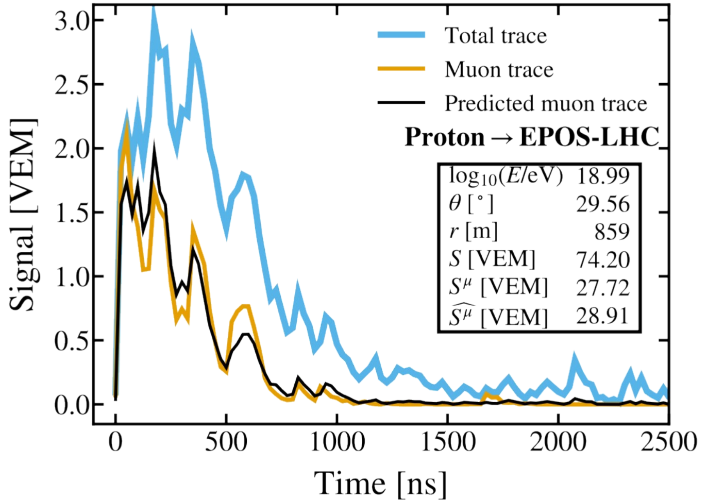
- Curve for the validation set decreases as the epoch increases
- Curve for validation is below the one for training because there are more events for which the performance is worse (lower zenith angles) in the training set
- Several tests had done to improve the performance but not significantly changed
    - loss function (3 PMT signals)
    - neural network layers

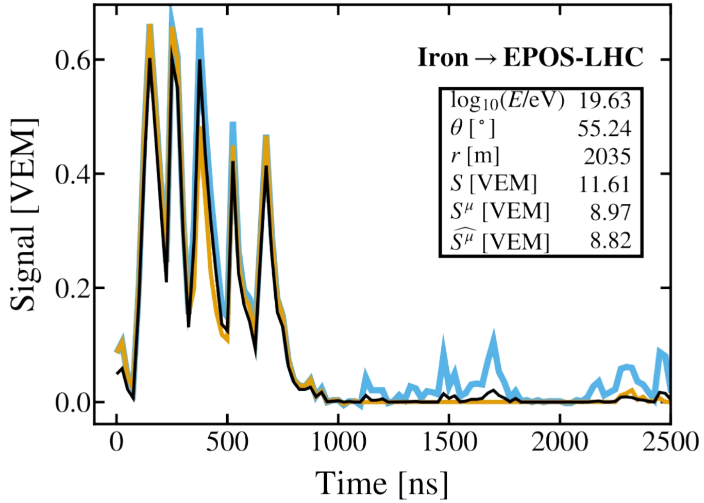




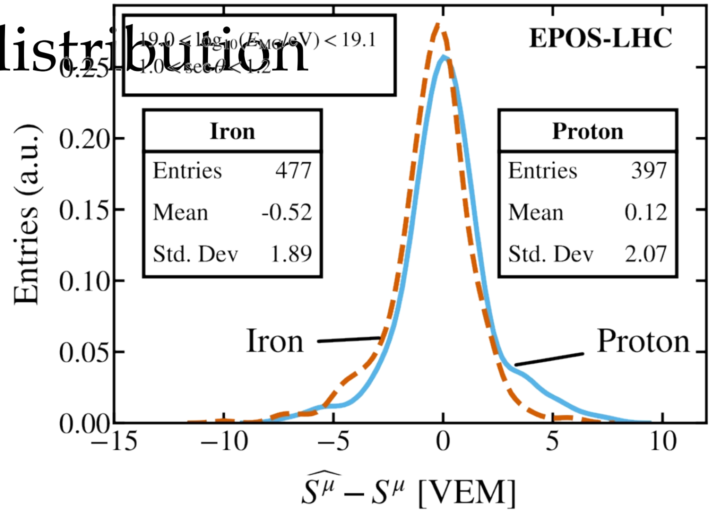
Results

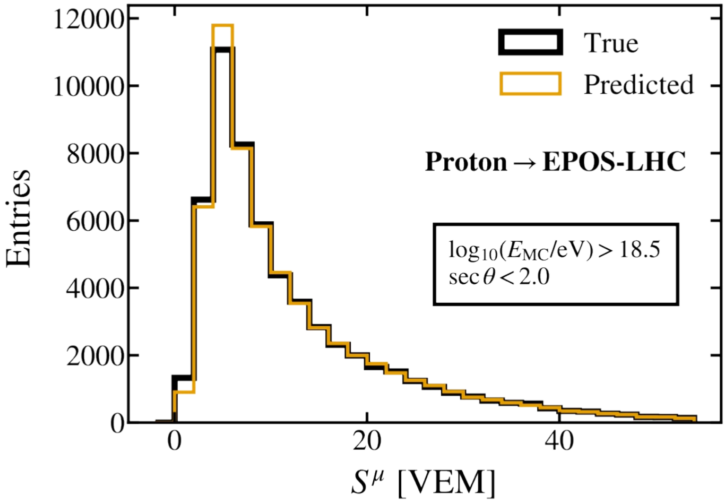
- Predicted muon traces examples





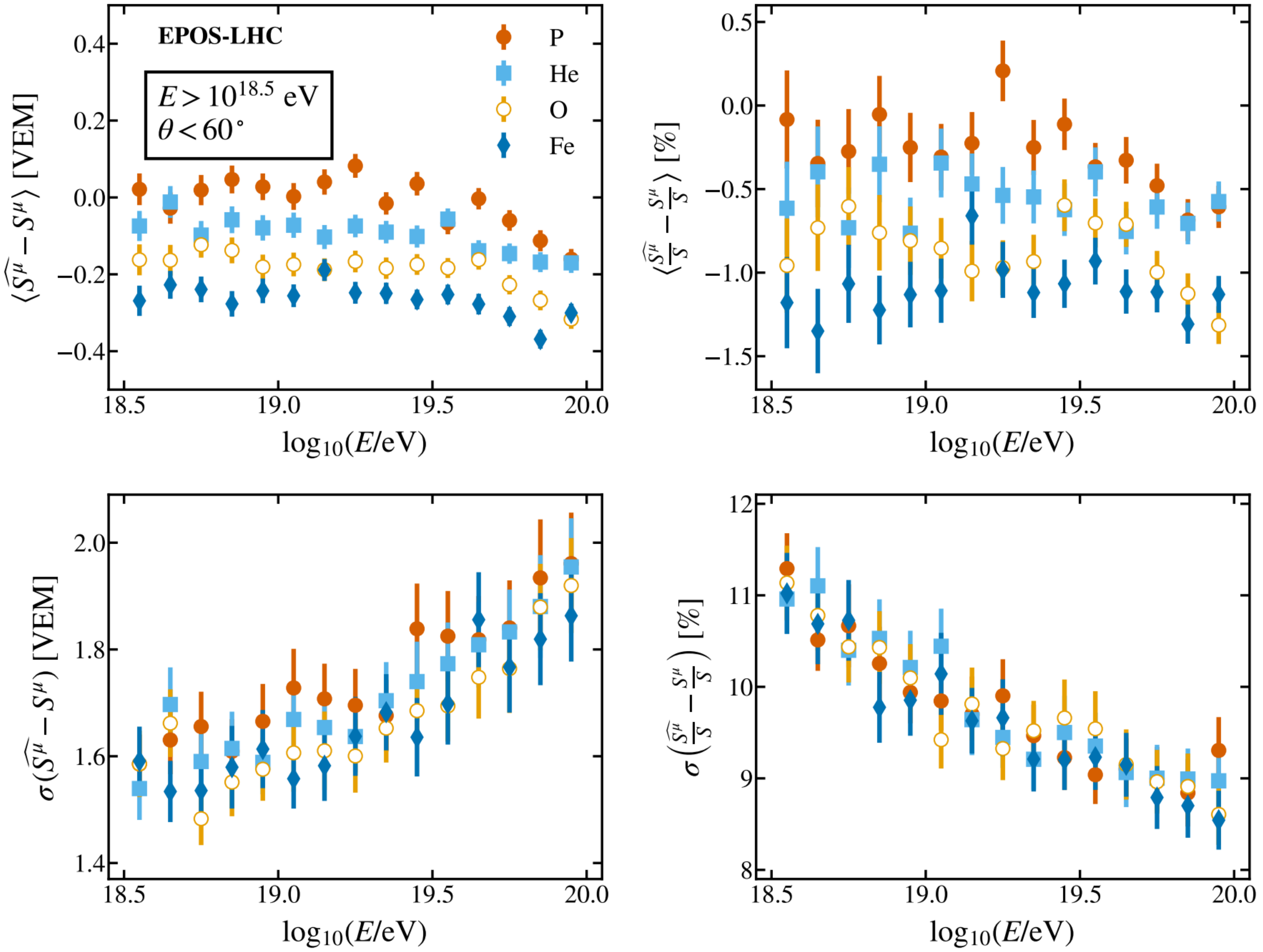
- $\boldsymbol{\Delta S^\mu}, \boldsymbol{\widehat{S_i^\mu}}$ and $\boldsymbol{S_i^\mu}$ distribution





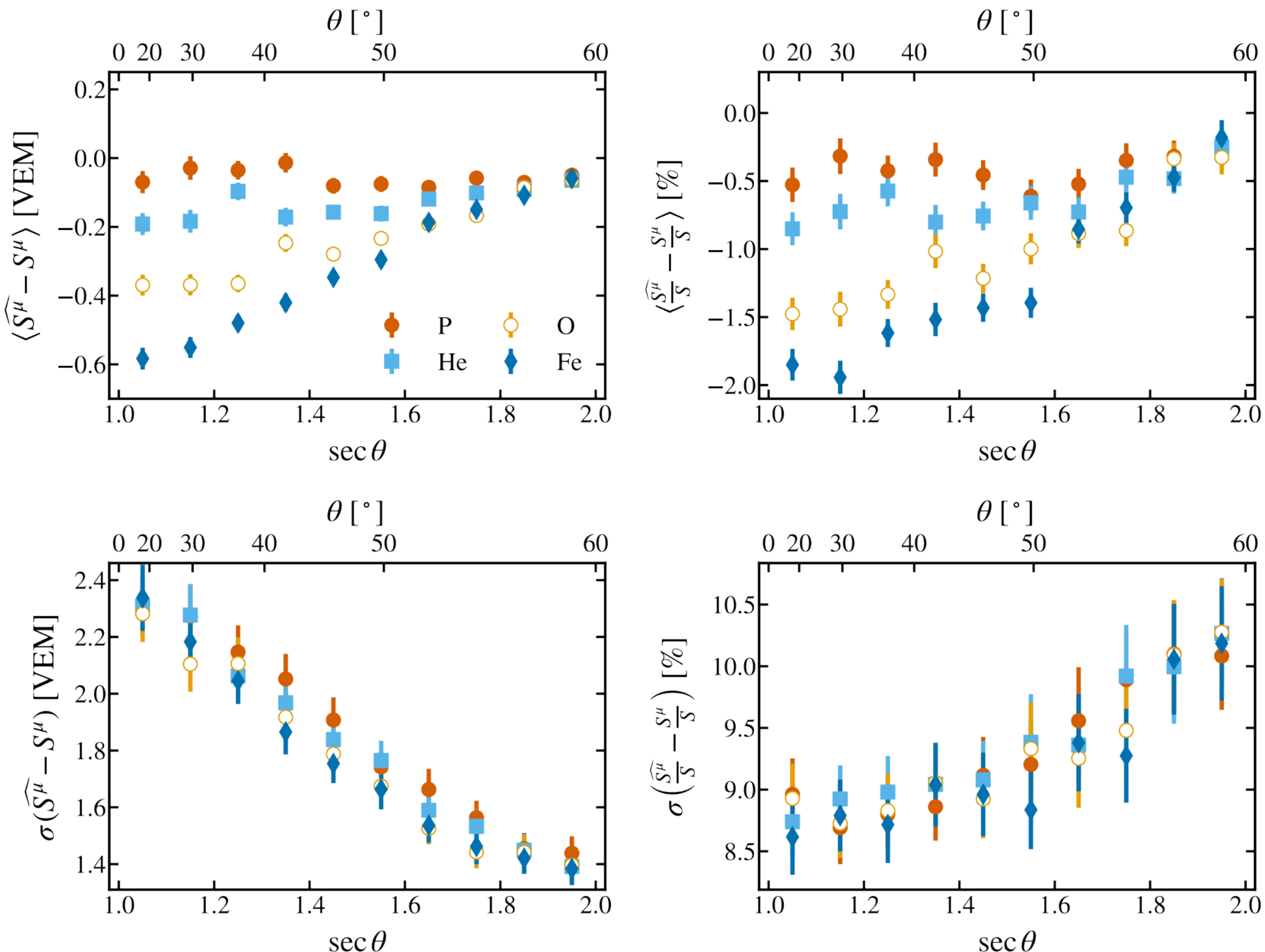
Results

- Bias and resolution as a function of energy



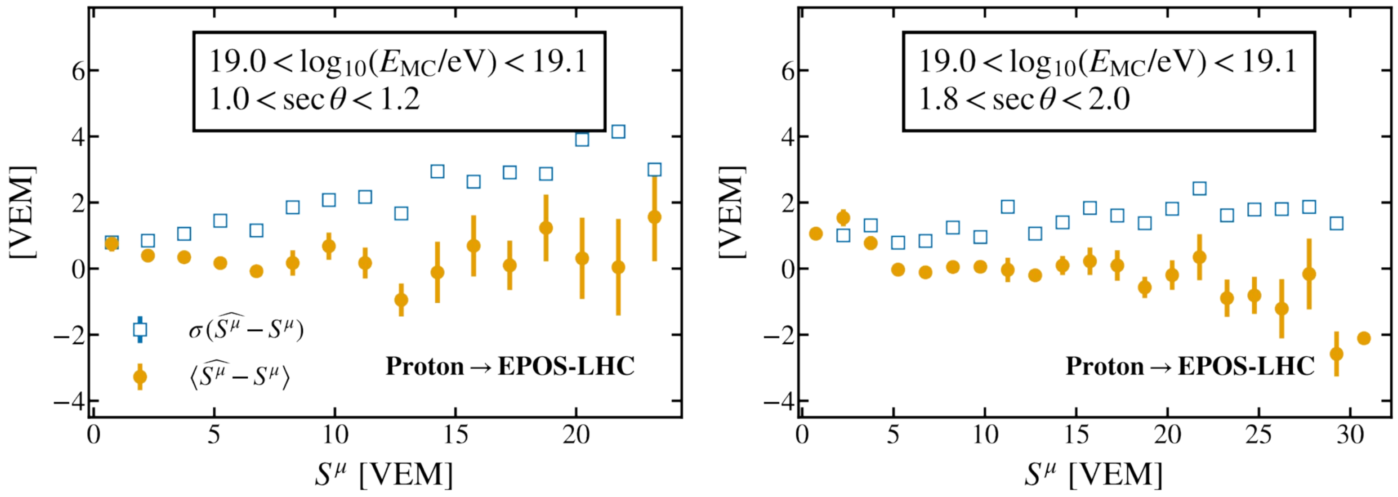
Results

- Bias and resolution as a function of zenith angle

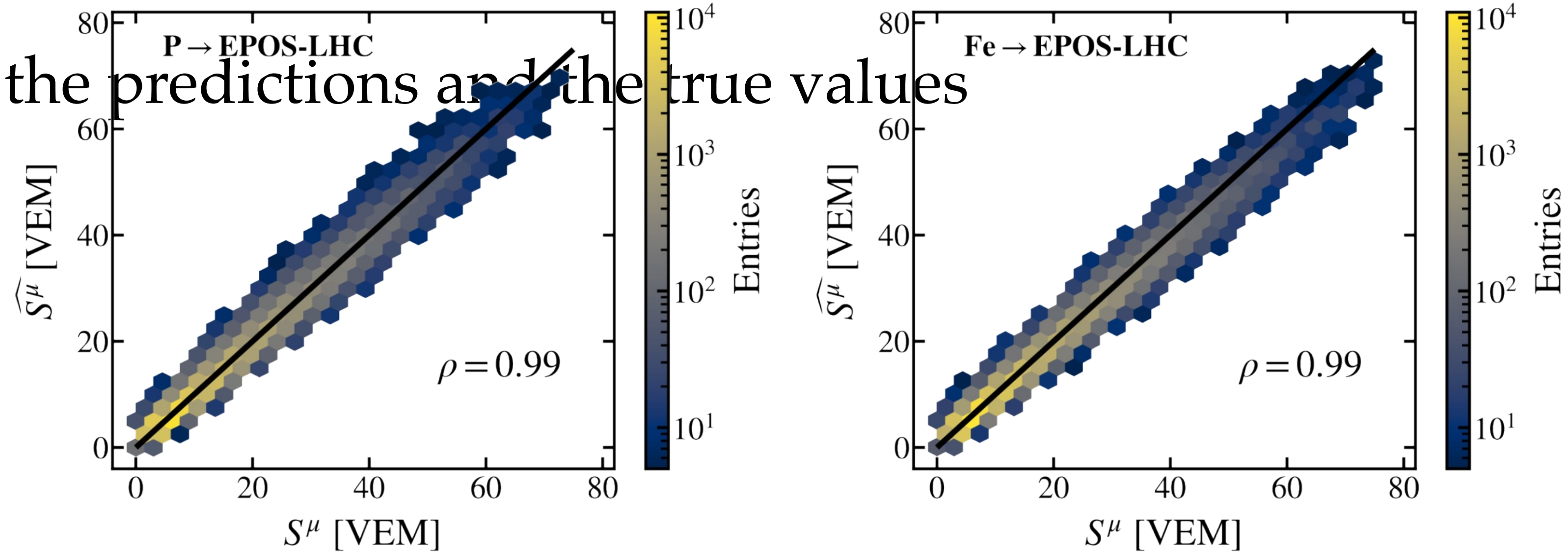


Results

- Bias and resolution as a function of $\boldsymbol{\mu}$ signal size

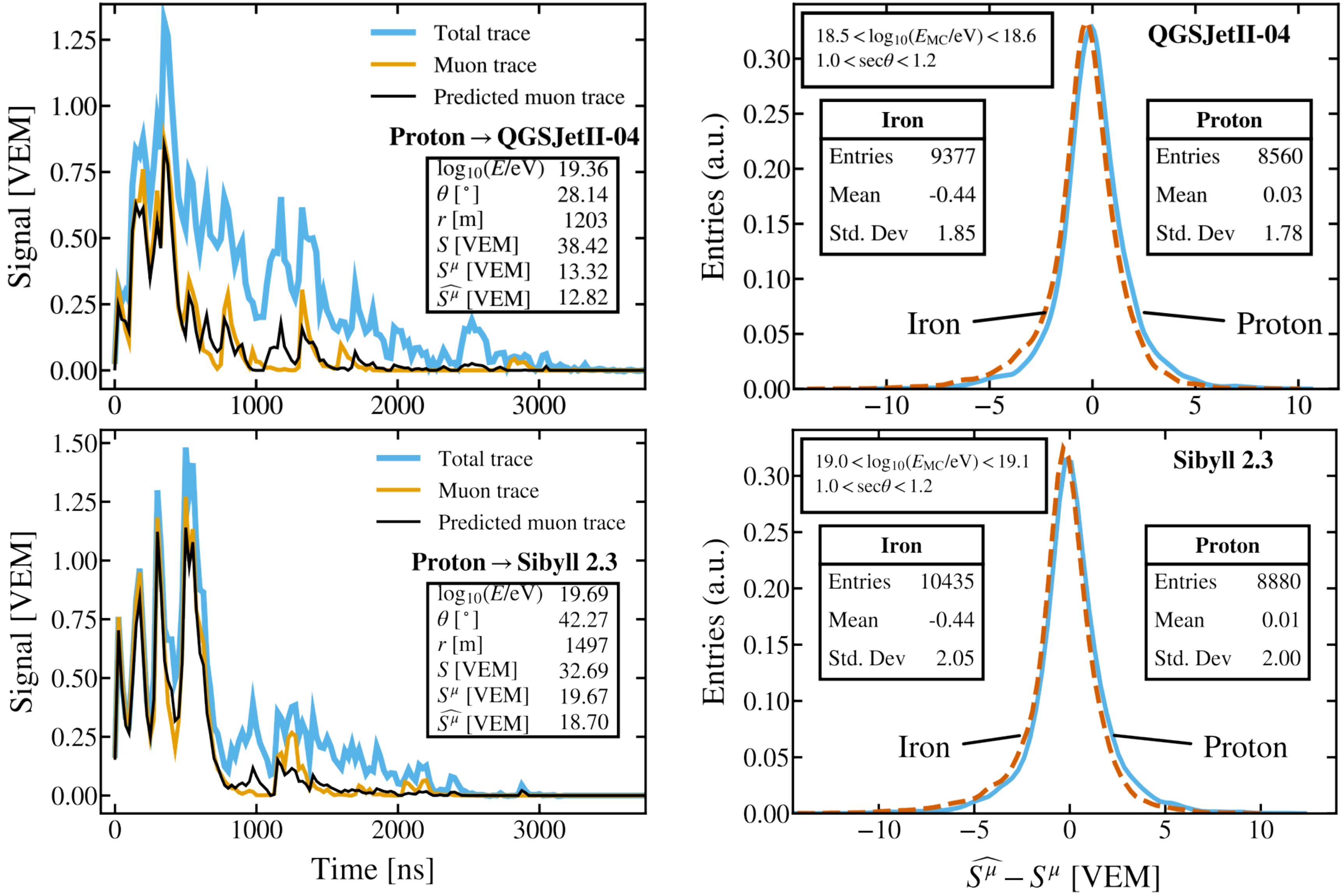


- density plot of the predictions and the true values



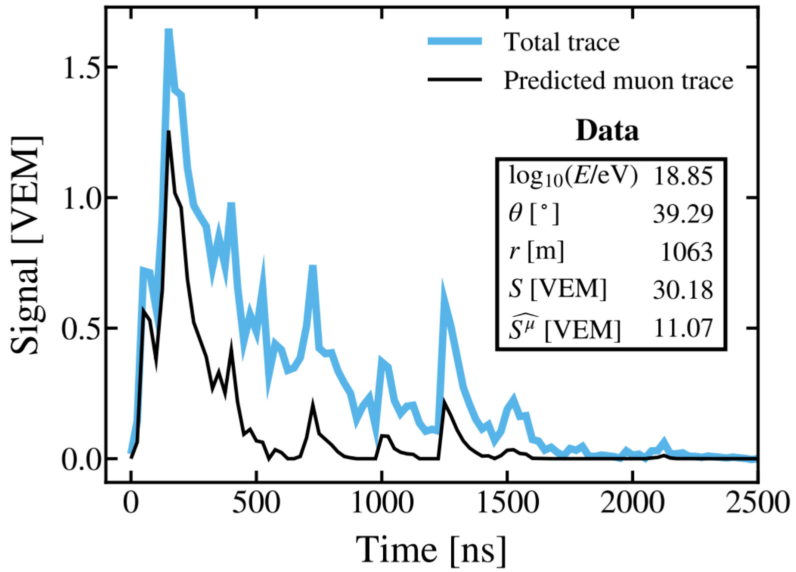
Results

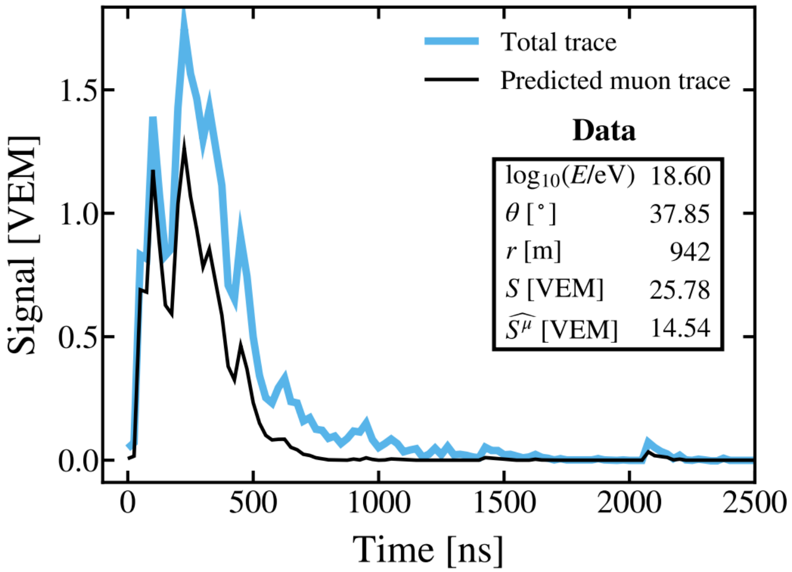
- Hadronic interaction model



Results

- Apply to data





Results

- Lateral distribution for $\boldsymbol{\mu}$ / electromagnetic component comparison with Akeno function
  - muon:

$$\rho_\mu = N_\mu\left(C'_\mu/R_0^2\right)r^{-0.75}(1+r)^\beta\left\{1+(R/800\mathrm{m})^3\right\}^{-\delta}$$

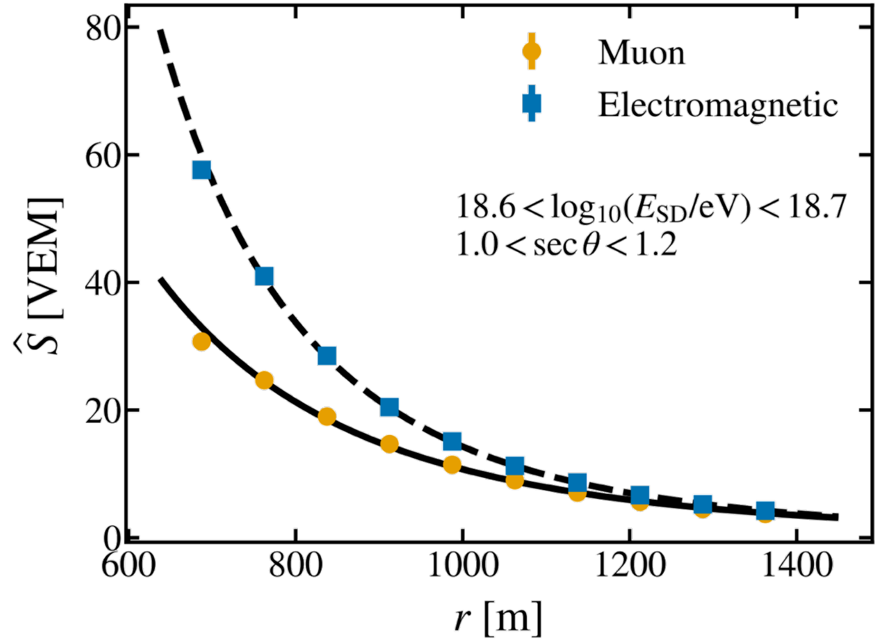
$$\log(R_0) = (0.58 \pm 0.04)(\sec\theta - 1) + 2.39 \pm 0.05$$

$$C'_\mu = 0.325, \beta = 2.52, \delta \sim 0.6$$

  - electromagnetic:

$$\rho_e = N_e C_e R^{-\alpha}(1+R)^{-(-\eta-\alpha)}\left(1+\frac{r}{2000}\right)^{-0.5}$$

$$R = r/R_M, \alpha = 1.2$$



10^{19} eV study
(for backup)

μ MC dataset, inputs for RNN

- CORSIKA
  - interaction model: QGSJetII-04
  - primary: proton/iron
  - energy range: 10^{19} eV only
  - zenith angle: 0 - 70 deg
- detector simulation with TA SD array
- TA SD standard reconstruction had performed (reconstructed by sdanalysis)
  - should be passed TASD spectrum cuts except for zenith angle cut
    - $\boldsymbol{\theta}_{\mathbf{rec}} < 60$deg applied
  - signal selection
    - NOT saturated
    - signal > 1.5 VEM
  - first waveform only (total ~`600000` waveforms for training)

this dataset ONLY USE E=10EeV

- Inputs
  - sec$\boldsymbol{\theta}$, distance from core (NOT lateral distance)
  - 128 bins waveform, both layers

Data selection and training

- distance $\boldsymbol{r}$ and $\mathbf{sec}\boldsymbol{\theta}$ are scaled to be between 0 and 1
- all traces are scaled individually to be between 0 and 1
- True (simulated) muon trace is scaled with the same factor used for the total trace
    - both **upper/lower trace using (NOT summation), 256 datapoints in total**

- Output of the neural network is also between 0 and 1 so as to use the same factor to rescale back the predicted trace

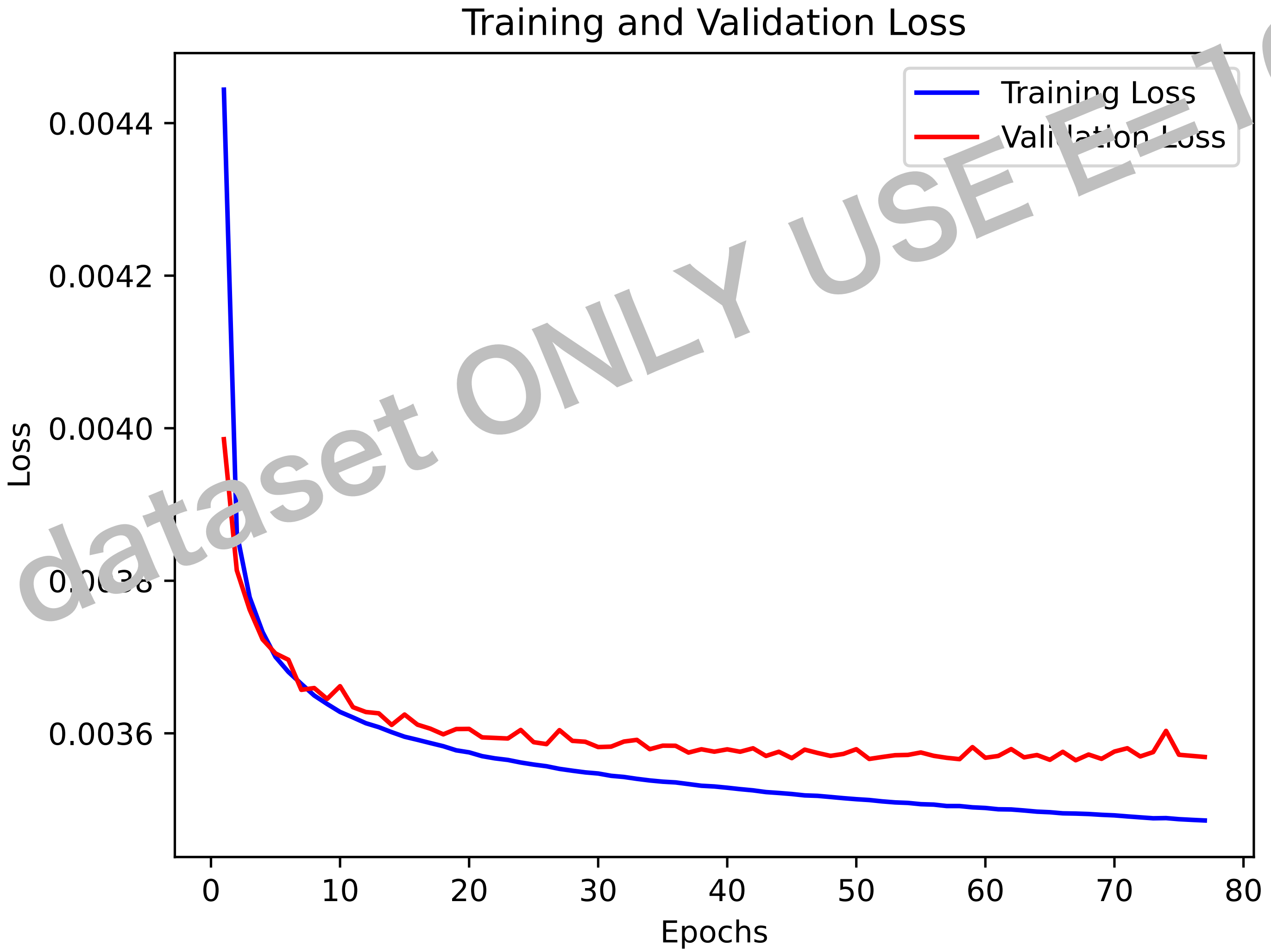
- loss function: $$L = \frac{1}{128\times 2}\sum_{i=1}^{128\times 2}\left(\widehat{S_i^{\mu}} - S_i^{\mu}\right)^2$$

- optimizer: ADAM

- Nvidia 3070 Ti GPU, training takes around 0.5-1 hours for each model

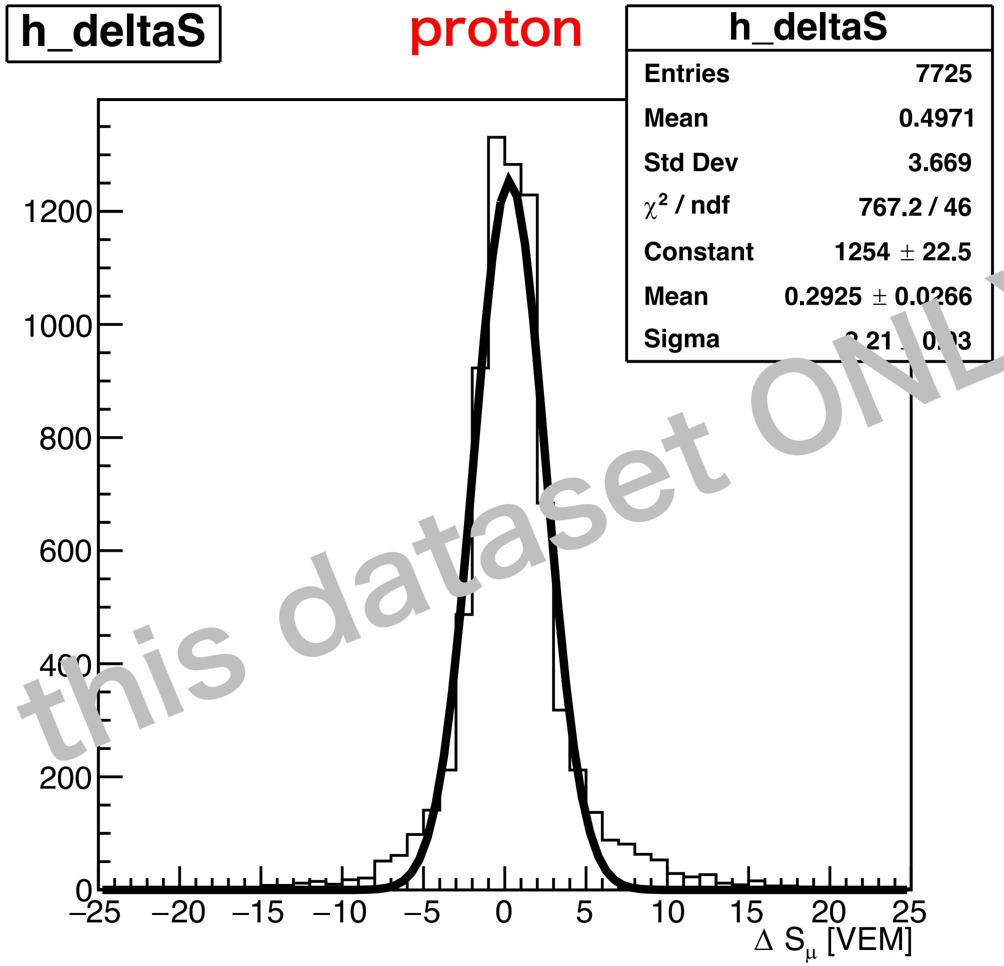
Result

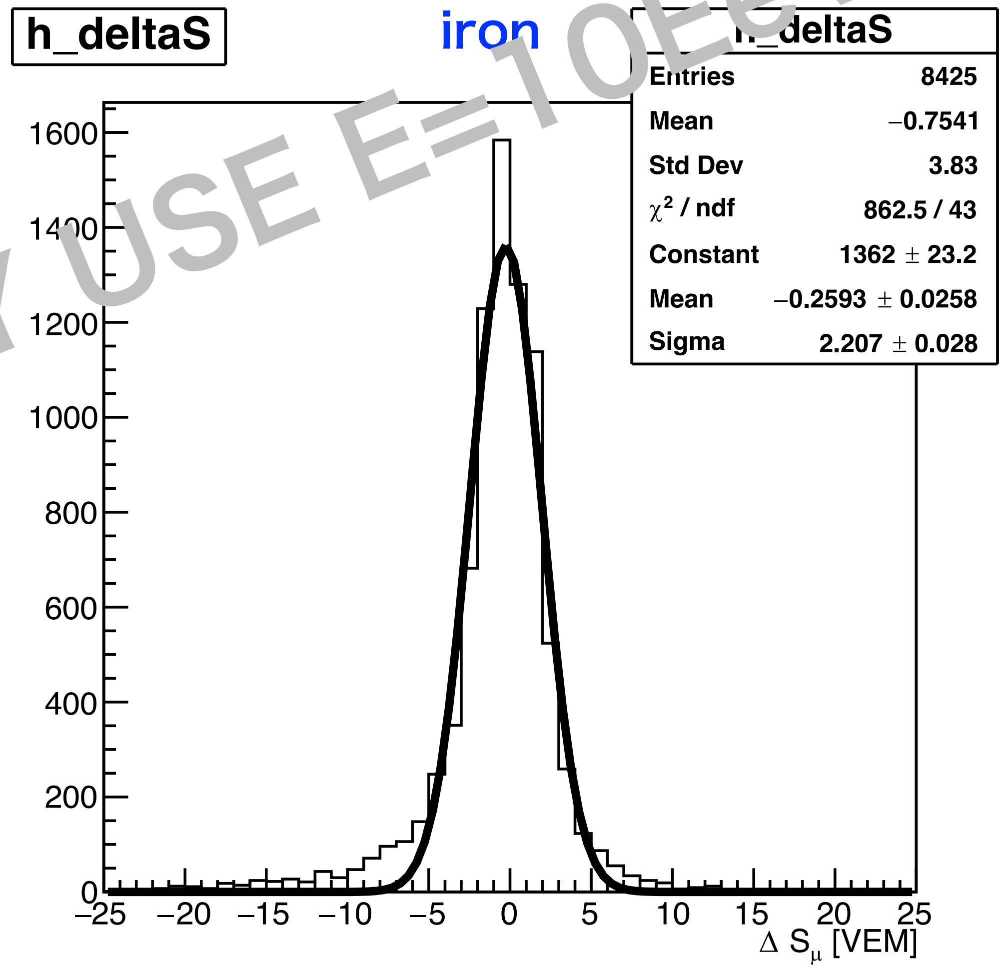
- loss


Training and Validation Loss
Training Loss
Validation Loss
0.0044
0.0042
0.0040
0.0038
0.0036
Loss
0
10
20
30
40
50
60
70
80
Epochs

Result

- muon prediction resolution: $\boldsymbol{\Delta S_{\mu} = S_{\mu,\mathrm{pre}} - S_{\mu,\mathrm{true}}}$ [VEM]

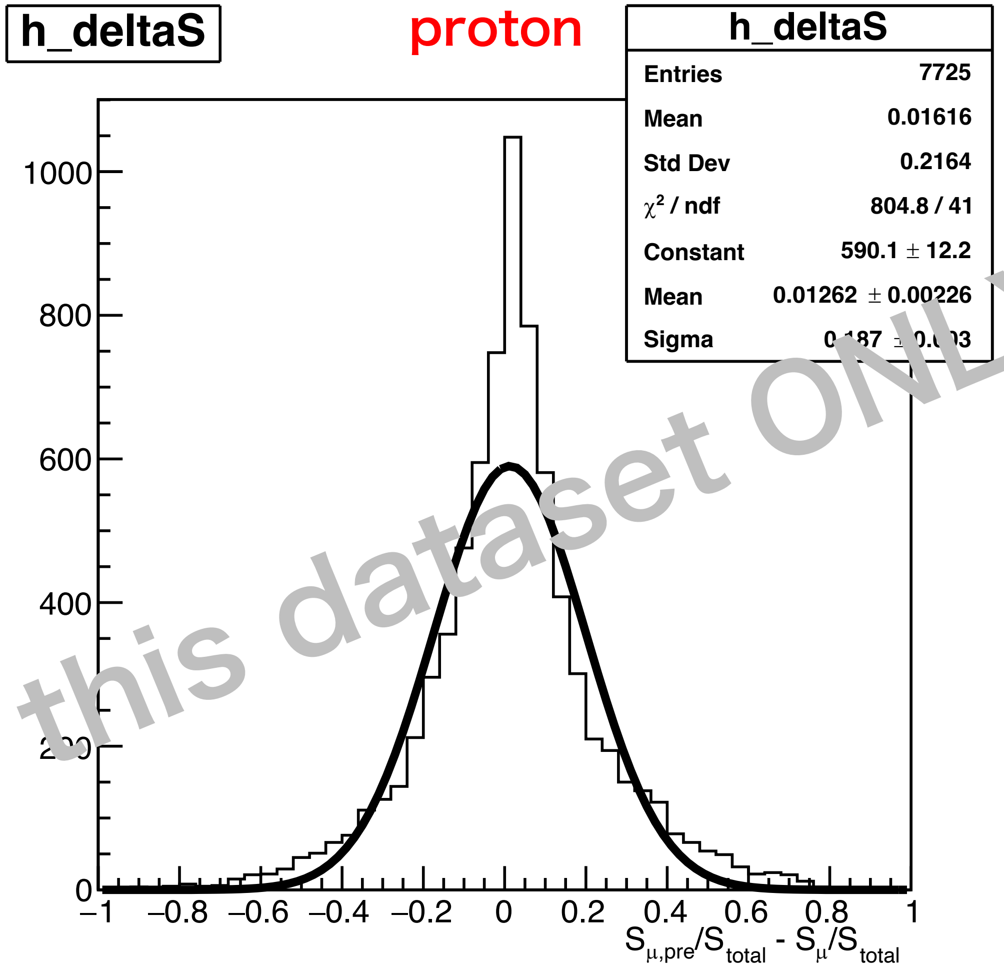


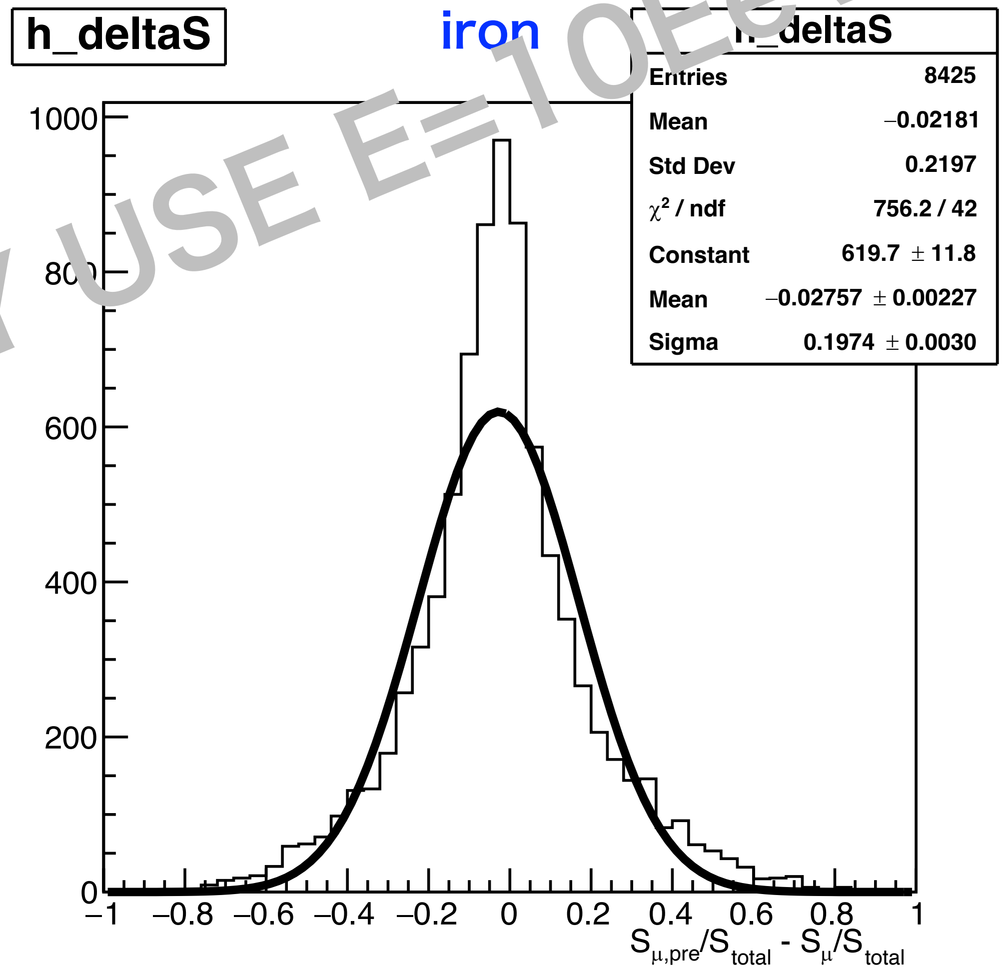


Result

- muon prediction resolution: $S_{\mu,\mathrm{pre}}/S_{\mathrm{total}} - S_{\mu,\mathrm{true}}/S_{\mathrm{total}}$ [%]

this dataset ONLY USE E=10EeV

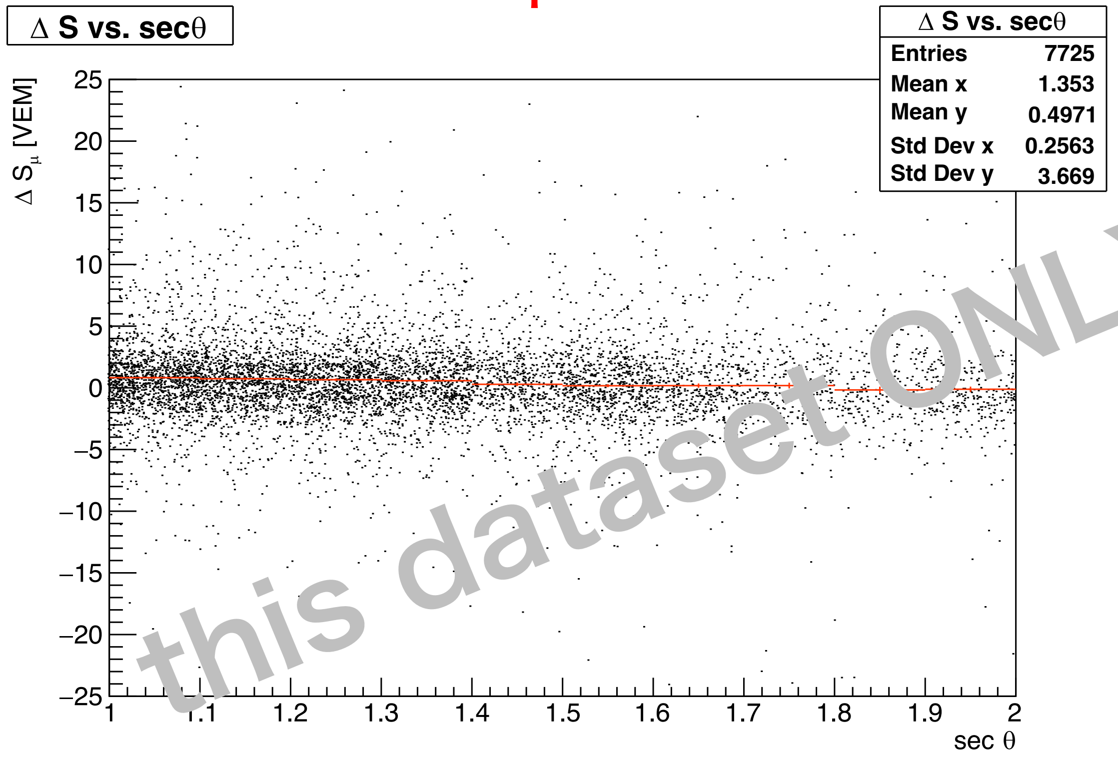




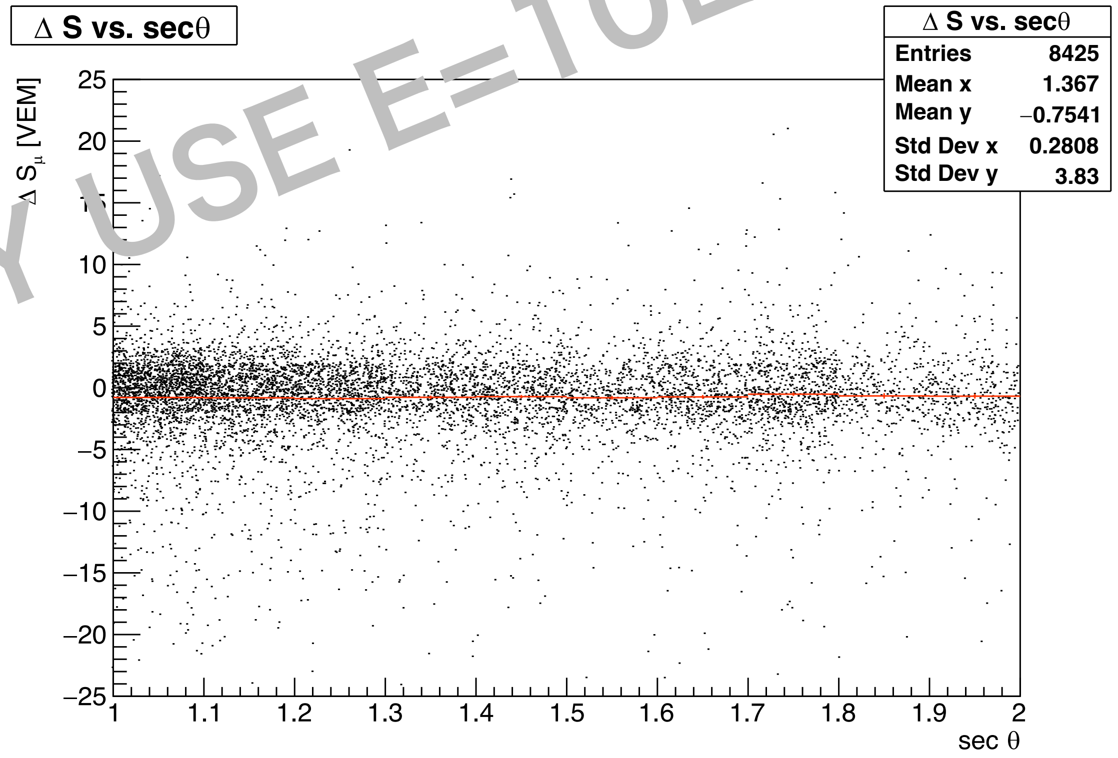
Result

- muon prediction, ΔS_{μ} vs. $\sec\theta$
  - No $\sec\theta$ dependence

proton



iron



Deep-learning based reconstruction of the shower maximum $X_{\mathbf{max}}$ using the water-Cherenkov detectors of the Pierre Auger Observatory

Pierre Auger collaboration

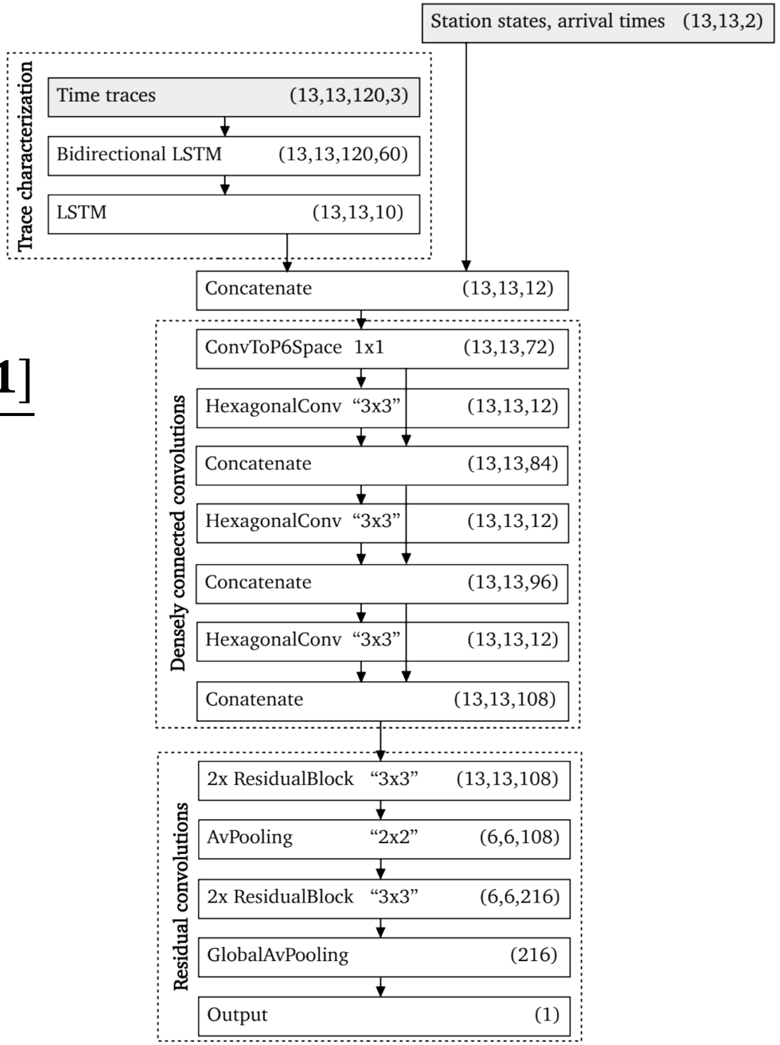
Abstract:
The atmospheric depth of the air shower maximum Xmax is an observable commonly used for the determination of the nuclear mass composition of ultra-high energy cosmic rays. Direct measurements of Xmax are performed using observations of the longitudinal shower development with fluorescence telescopes. At the same time, several methods have been proposed for an indirect estimation of Xmax from the characteristics of the shower particles registered with surface detector arrays. In this paper, we present a deep neural network (DNN) for the estimation of Xmax. The reconstruction relies on the signals induced by shower particles in the ground based water-Cherenkov detectors of the Pierre Auger Observatory. The network architecture features recurrent long shortterm memory layers to process the temporal structure of signals and hexagonal convolutions to exploit the symmetry of the surface detector array. We evaluate the performance of the network using air showers simulated with three different hadronic interaction models. Thereafter, we account for long-term detector effects and calibrate the reconstructed Xmax using fluorescence measurements. Finally, we show that the event-by-event resolution in the reconstruction of the shower maximum improves with increasing shower energy and reaches less than 25 g/cm2 at energies above 2×10^{19} eV.

doi: 10.1088/1748-0221/16/07/P07016

Architecture

- Inputs
  - 13 x 13 WCDs around the WCD with the highest signal
    - 13 x 13 hexagonal inputs
      - Particle arrival times
      - Station states (1 or 0)
    - 3PMT signal traces into RNNs
      - first 3μs signal trace
      - rescaled signal

$$\widetilde{S}_i(t) = \frac{\log_{10}[S_i(t)/\mathrm{VEM} + 1]}{\log_{10}[100 + 1]}$$



- Output
  - $\boldsymbol{X}_{\mathbf{max}}$

Data selection and training

- CORSIKA
  - interaction model: EPOS-LHC
  - primary: proton/helium/oxygen/iron
  - energy range: 1 - 160 EeV
  - zenith angle: 0 - 65 deg

- Only traces for which the integral > 5 VEM are selected
- 450 000 events were available
  - events were sampled randomly and assigned to the training, validation or test data sets using a uniform distribution in energy and secθ for the validation and test sets; the remaining events were assigned to the training set.
  - split as follows:
    400 000 in the training set
    ? in the validation set (including above??)
    50 000 in the test set

Data selection and training

- loss function:

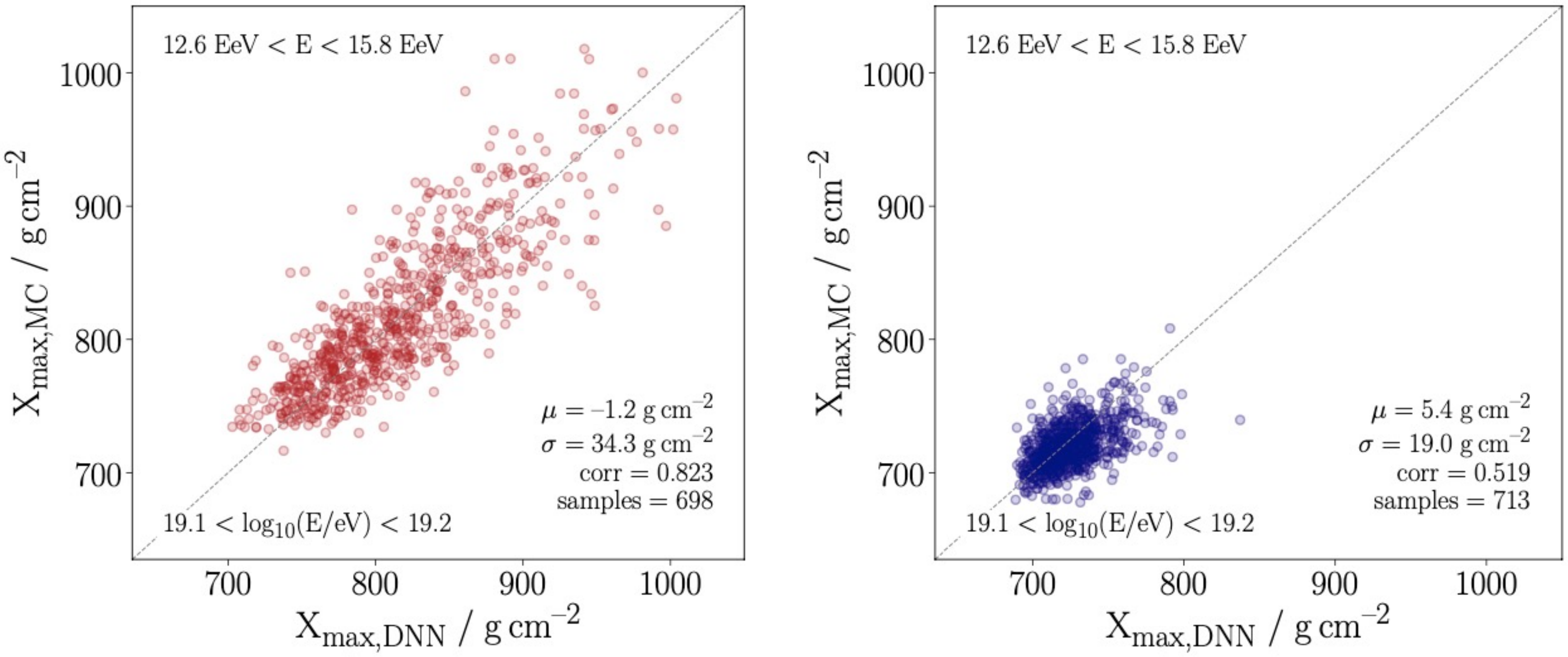
$$L = \sum_{Z=\mathrm{H,He,O,Fe}} \left(\hat{X}_{\mathrm{max},Z} - X_{\mathrm{max},Z}\right)^2$$

- optimizer: ADAM

- Nvidia GTX 1080 GPU, the training takes around 60 hours

Results

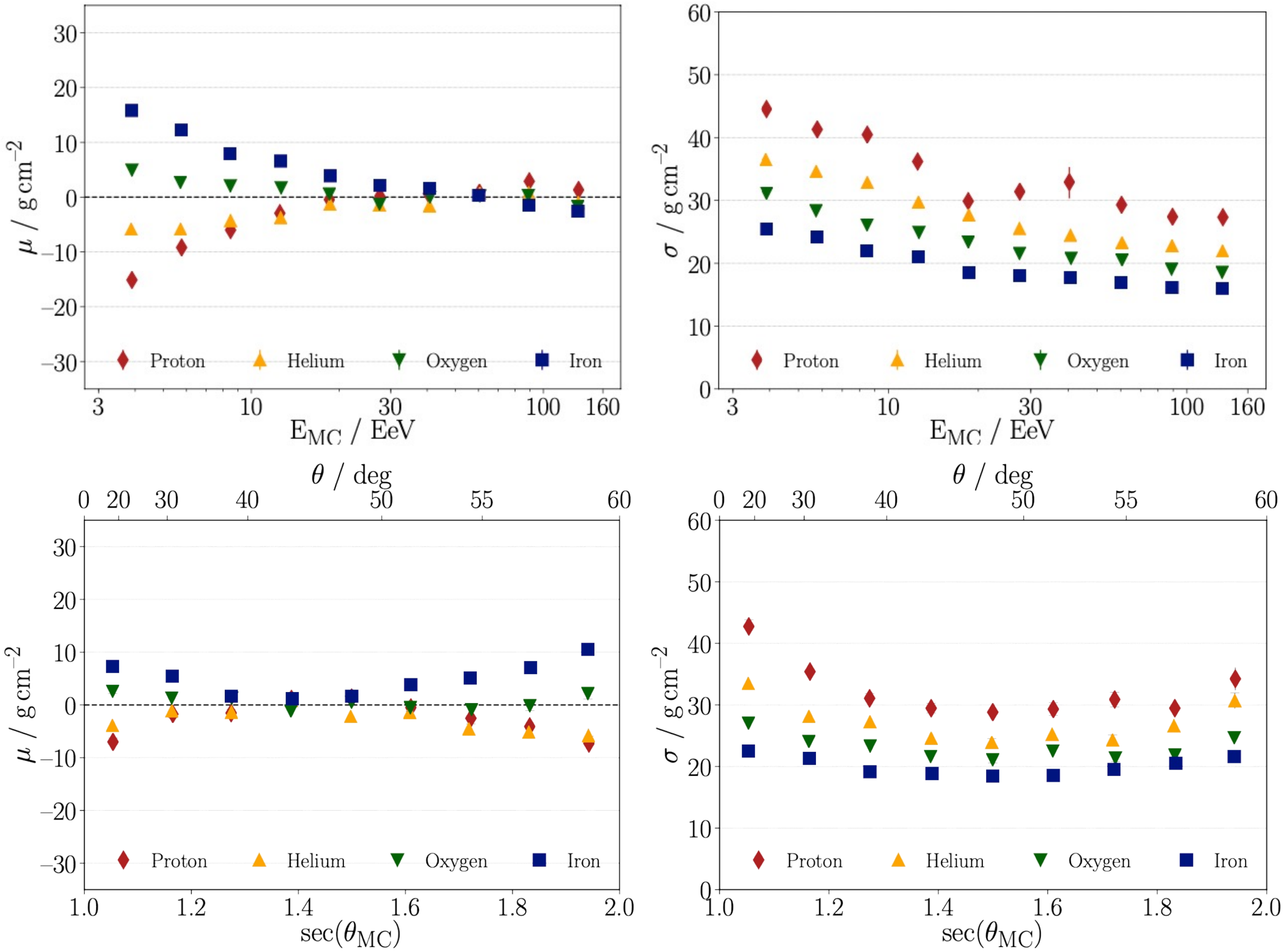
- Correlation of simulated and predicted $\boldsymbol{X}_{\mathbf{max}}$
  - $19.1 < \log(E/eV) < 19.2$



(a) Proton (b) Iron

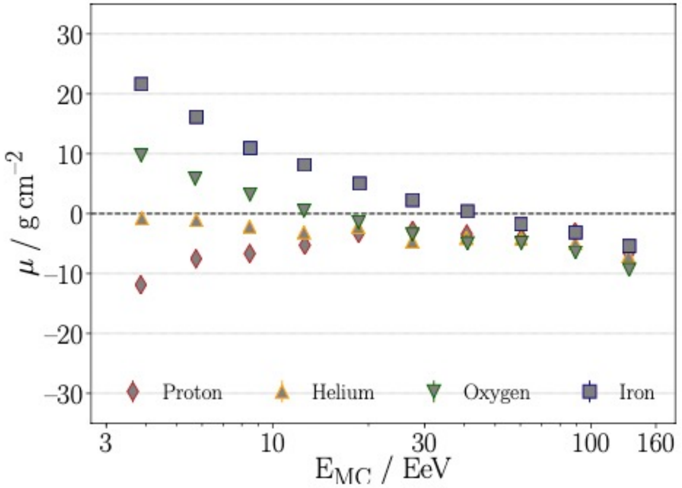
Results

- $X_{\mathbf{max}}$ bias and resolution

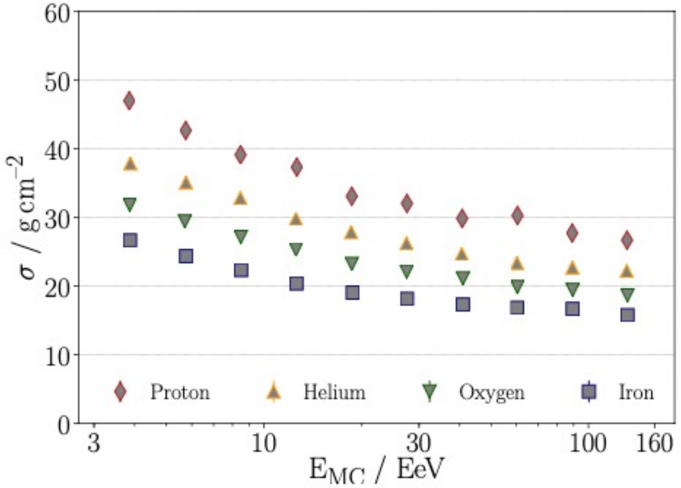


Results

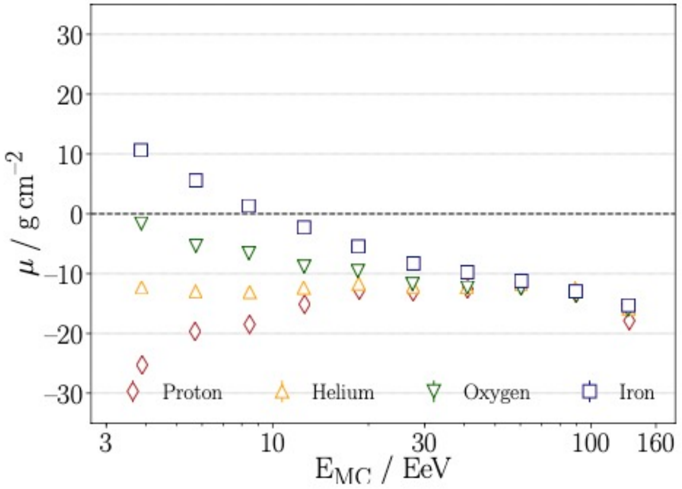
- interaction model dependences



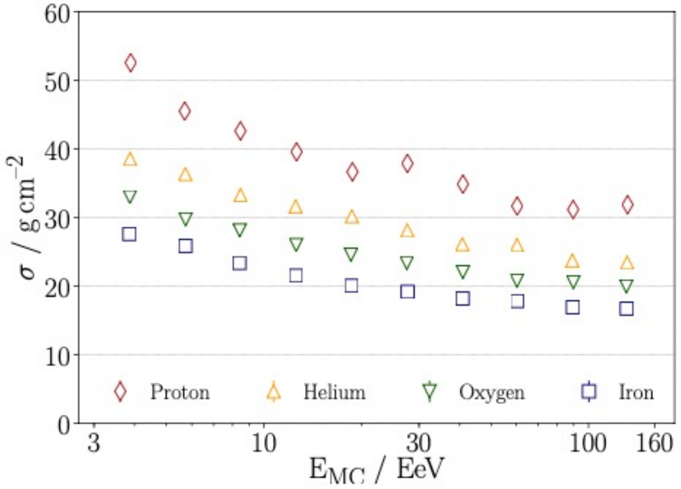
(a) QGSJetII-04: X_{max} bias



(b) QGSJetII-04: X_{max} resolution



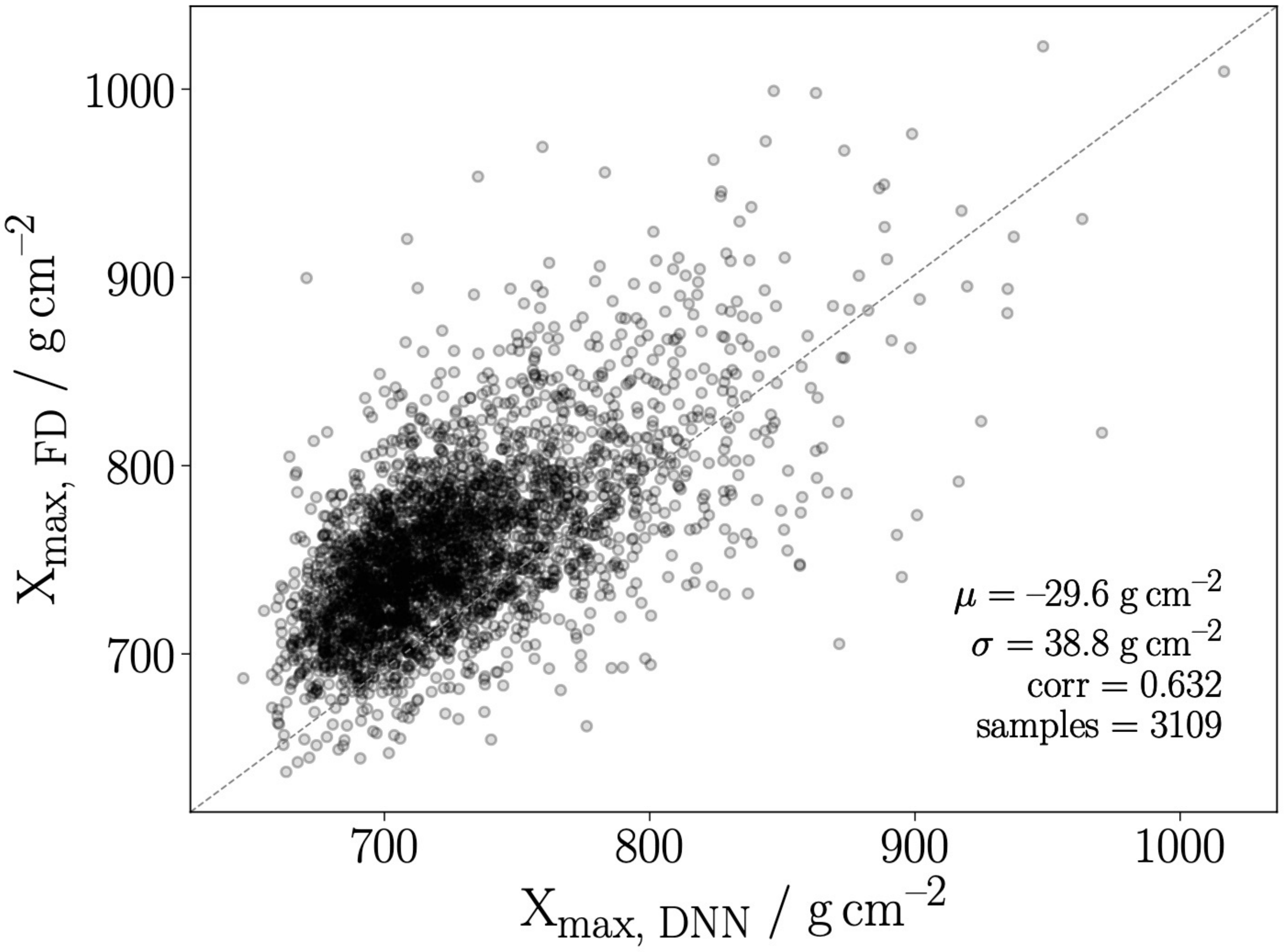
(c) Sibyll 2.3: X_{max} bias



(d) Sibyll 2.3: X_{max} resolution

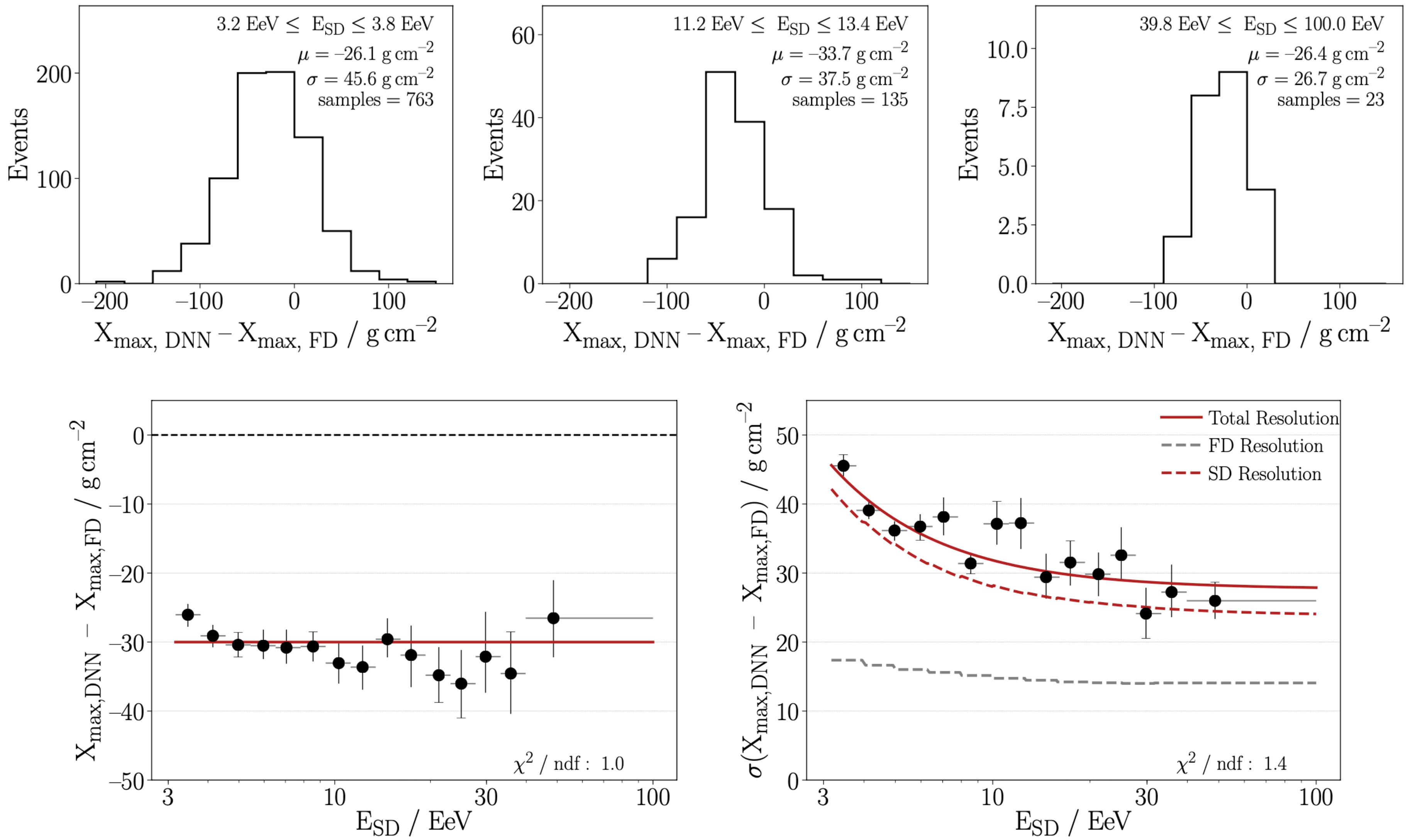
Results

- Apply to data $\boldsymbol{X}_{\mathbf{max,DNN}}$ vs. $\boldsymbol{X}_{\mathbf{max,FD}}$



Results

- Apply to data



Comments

- As shown in the page 12, apparently the upper layer waveform is retained in the lower layer unlike other components like electrons. So this analysis is so important firstly fully utilizing the data of two layers.
- Anatoli Fedynitch から 全員 へ (14:37)
- Yes, agree!
- 常定　芳基 から 全員 へ (14:37)
- If the output of the current implementation of RNN is the waveforms of the both layers, how about reducing the output to only the upper layer or the average of the two layers? Is it posible to improve the accuracy of muon numbers by doing that?
- 自分から 全員 へ (14:40)
- I didn’t do such kind of examinations, so I can’t say how much improve the accuracy.
- But I can do that. I will show that next time
- 常定　芳基 から 全員 へ (14:41)
- 上層と下層の波形を両方予想するより、上層だけ、または上層下層の平均を予想のように出力を減らしたら精度が上がらないかなと思いました。いずれにしろ、２層データをフルに使う解析につながるので素晴らしい仕事と思います。
  - gain が違うチャンネルをどう平均する？ 最後の rescale factor が定まらん
  - 最初に一度 VEM(高いチャンネルの)でscaleして平均？
- 上層下層の波形が似ているというのはミューオン同定に大きなヒントのはずなので。