Mass composition of 1 - 100 PeV cosmic rays with KASCADE and machine learning

Nikita Petrov INR, BINP

Outline

- Goal & Motivation
- Experiment & Monte-Carlo
- Method
  - ML methods
  - Unfolding
- Results
- Conclusion

Goal

Measure the mass components spectra of cosmic rays (protons, He, C, Si, Fe) in the energy range 1 – 100 PeV

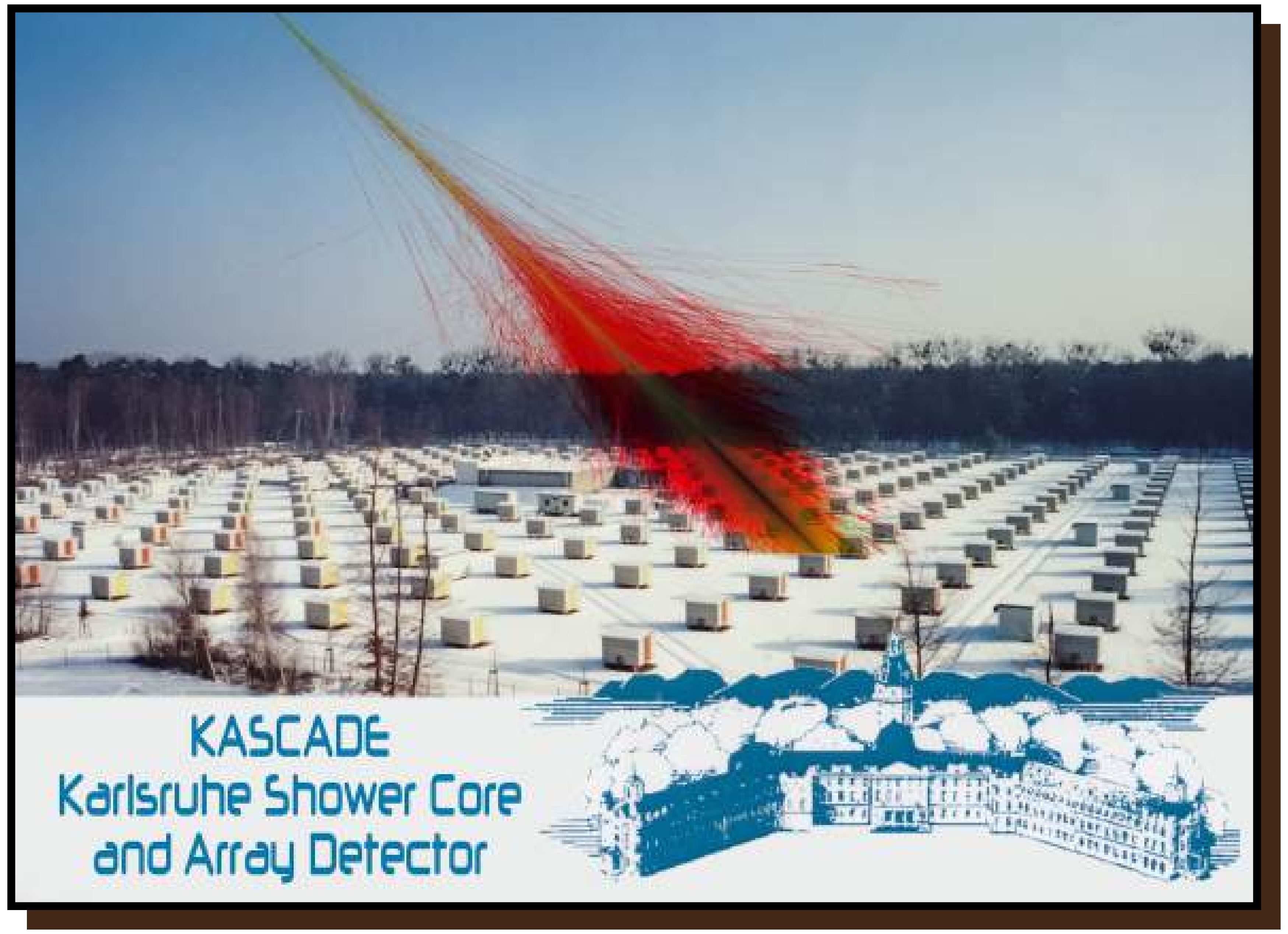
Motivation

- New machine learning methods of the analysis
- New hadronic interaction models
- Large amount of clean available data* comparable to the recent experiments
- Discrepancy between different experiments in the studied energy range

* Data are taken from KCDC: A.Haungs et al; Eur. Phys. J. C (2018) 78:741; "The KASCADE Cosmic ray Data Centre KCDC: granting open access to astroparticle physics research data"; (doi: 10.1140/epjc/s10052-018-6221-2)

Cosmic rays

- Cosmic rays are high-energy particles which arrive from space to the Earth's atmosphere
- Extensive air showers are cascades of subatomic particles and ionized nuclei, produced in the atmosphere when a primary cosmic ray enters the atmosphere.
- KASCADE detects cosmic rays via extensive air showers

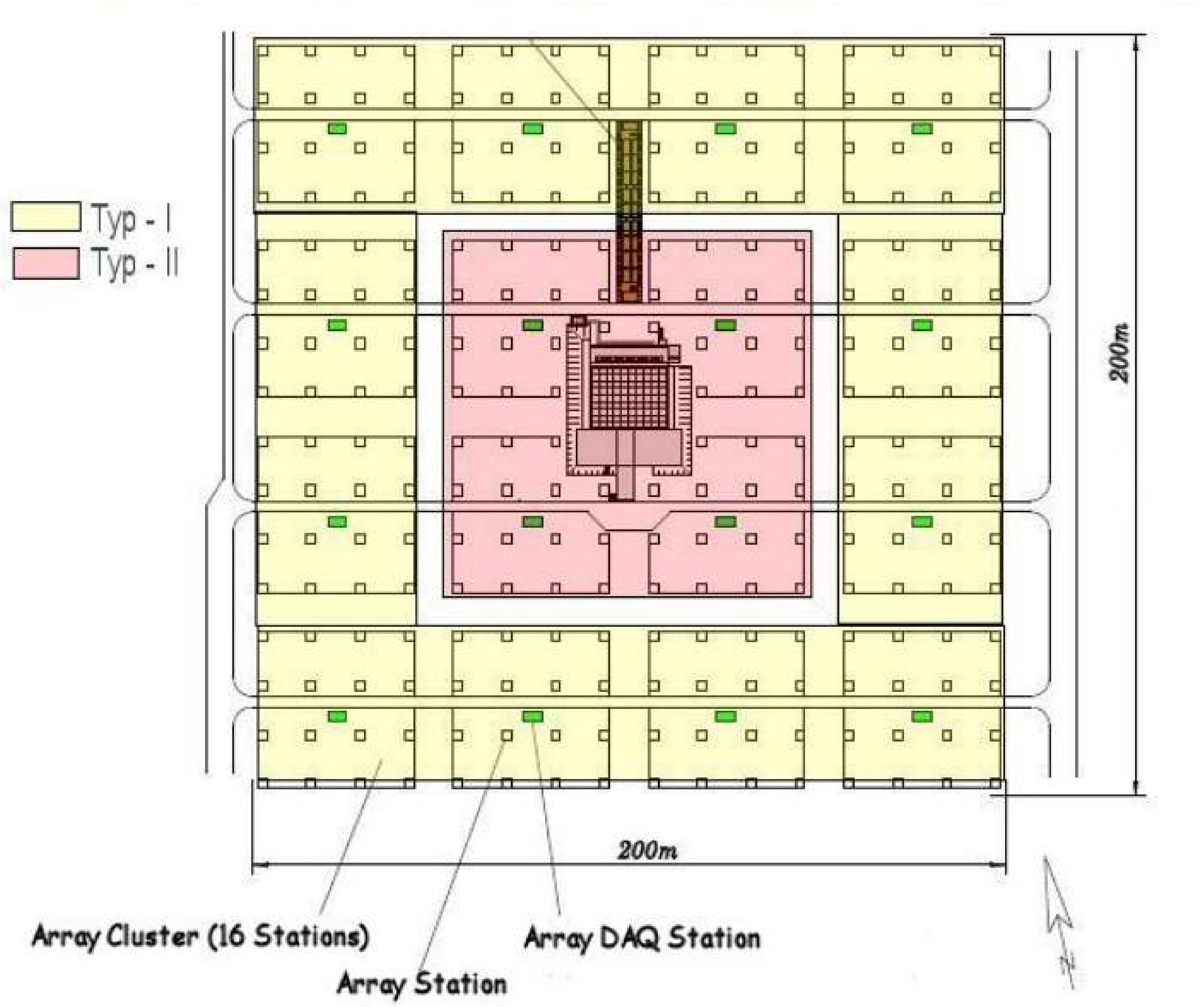


KASCADE experimental setup

KASCADE is an extensive air shower experiment that was located in KIT Campus, Karlsruhe, Germany (1996 - 2013)

KASCADE array: 252 scintillator detectors placed in rectangular grid with 13 m spacing and covering the area of 200 × 200 m² in total.

Energy range: ~ 500 TeV — 100 PeV



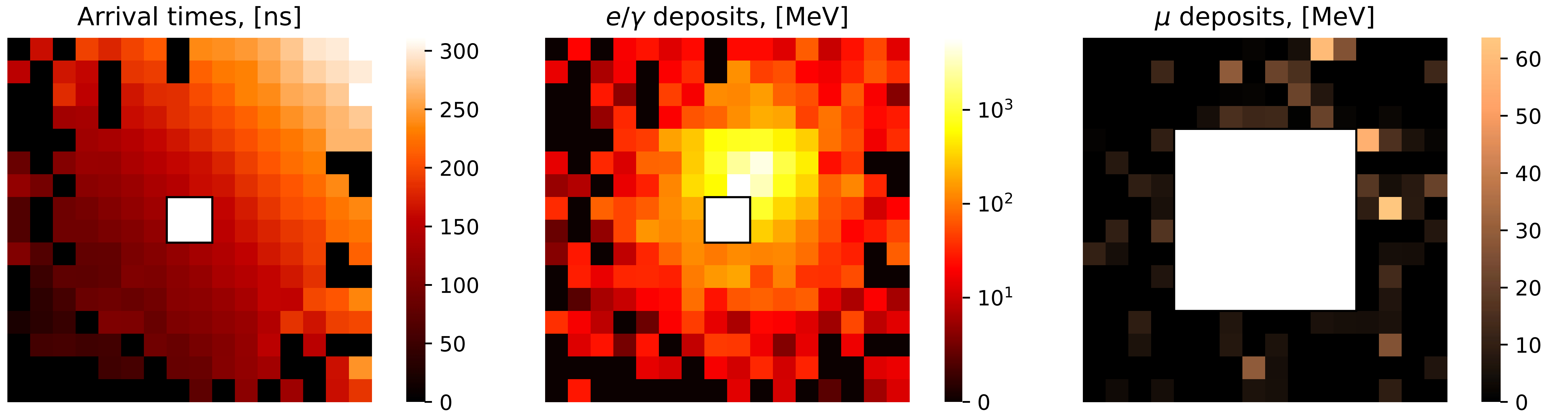
Schematic view of the KASCADE detector array (type I – e/γ and μ detectors, type II – e/γ only)

Experimental data and MC

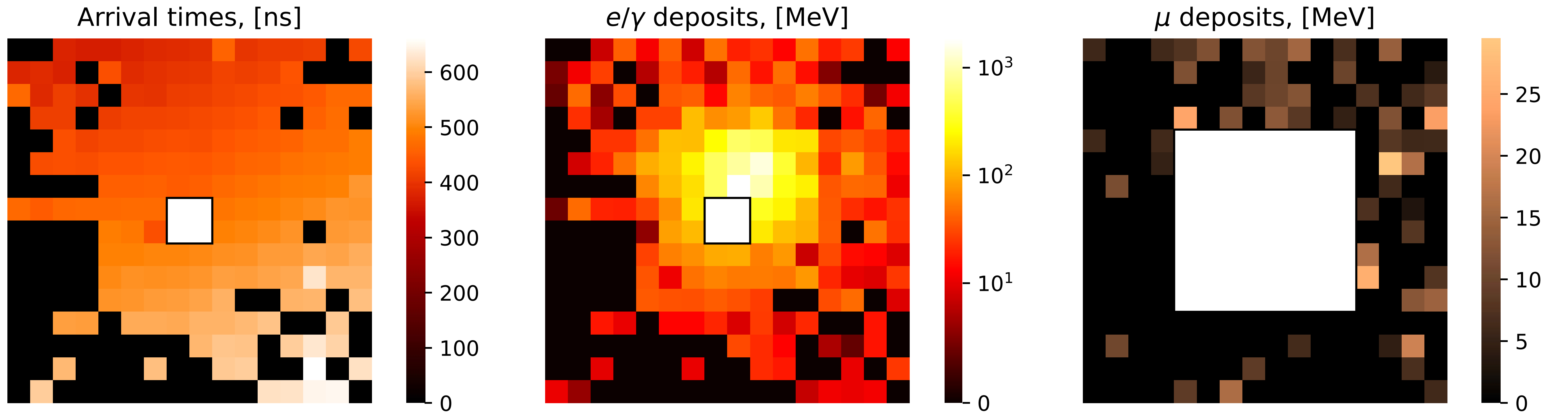
Event structure:

- 3 arrays, 16x16 shape
  - arrival times, [ns]
  - e/γ energy deposits, [MeV]
  - μ energy deposits, [MeV]
- reconstructed features
  - energy (E), zenith (θ) and azimuth angles (φ) of the primary particle
  - shower core position (x, y)
  - electron and muon total numbers (Ne, Nμ)
  - shower age parameter (s)

MC event: $\log_{10}$ E, [eV] = 15.51, θ = 20.78°



Exp event: $\log_{10}$ E, [eV] = 15.45, θ = 19.37°



Examples of the experimental and MC events from dataset. Missing detectors in the center region are shown in white.

Datasets and quality cuts

$\theta < 18°$
$\log_{10} N_e > 4.8$
$\log_{10} N_\mu > 3.6$
$x^2 + y^2 < 91$ m
$0.2 < s < 1.48$

Quality cuts applied to data and MC

Experiment dataset

Unblind: 20%	Blind: 80%

~ 8.5M events in total (after quality cuts)

Monte-Carlo datasets (CORSIKA + detector simulation)

QGSJet-II.04 ~ 180k events

EPOS-LHC | Sibyll 2.3c | QGSJet-II.02

Machine learning methods

Train different classifiers of the primary particle and select the best one

Random Forest (RF)

- as simple as possible
- input: x, y, E, Ne, Nμ, θ, φ, s

Convolutional NN (CNN)

- simple architecture (~ 30k parameters)
- input: detector deposits + Ne, Nμ, θ, s

Multi-layer perceptron (MLP)

- exploits the spatial-specific info
- input: detector deposits + θ, φ

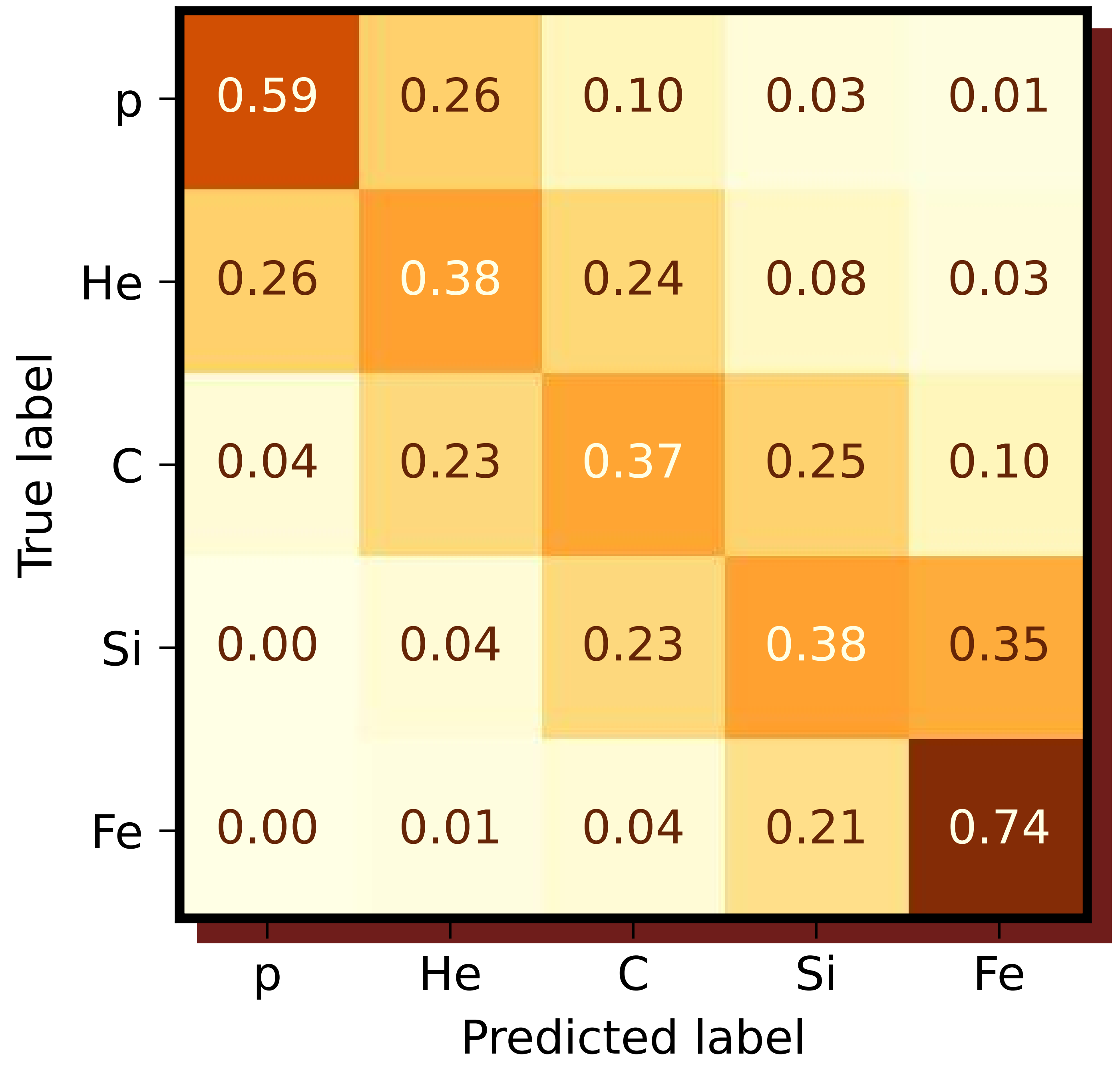
EfficientNet (EffNet)

- common standard architecture
- input: detector deposits + θ, φ

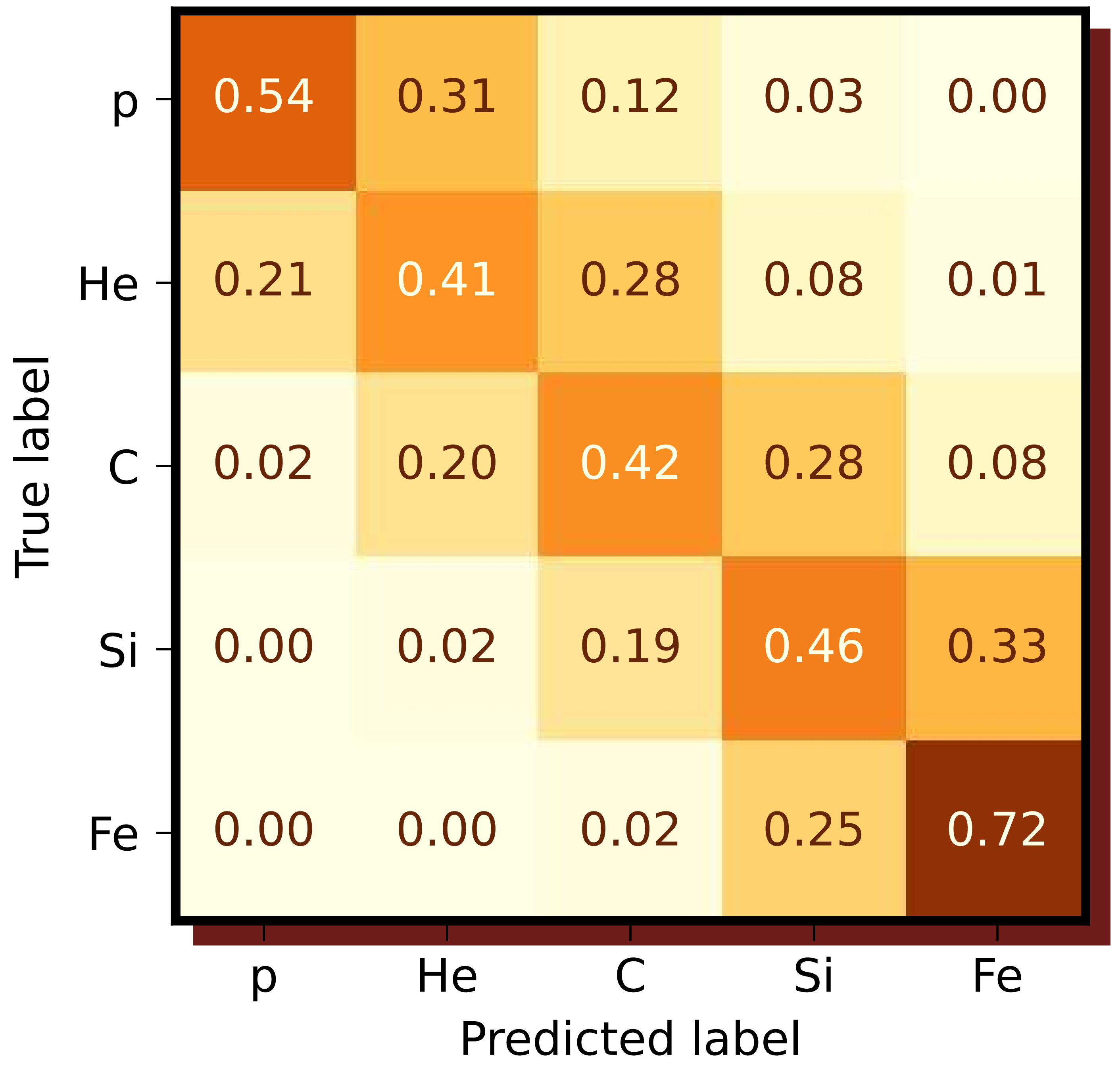
Performance of the ML methods

Compare confusion matrices for the applied methods

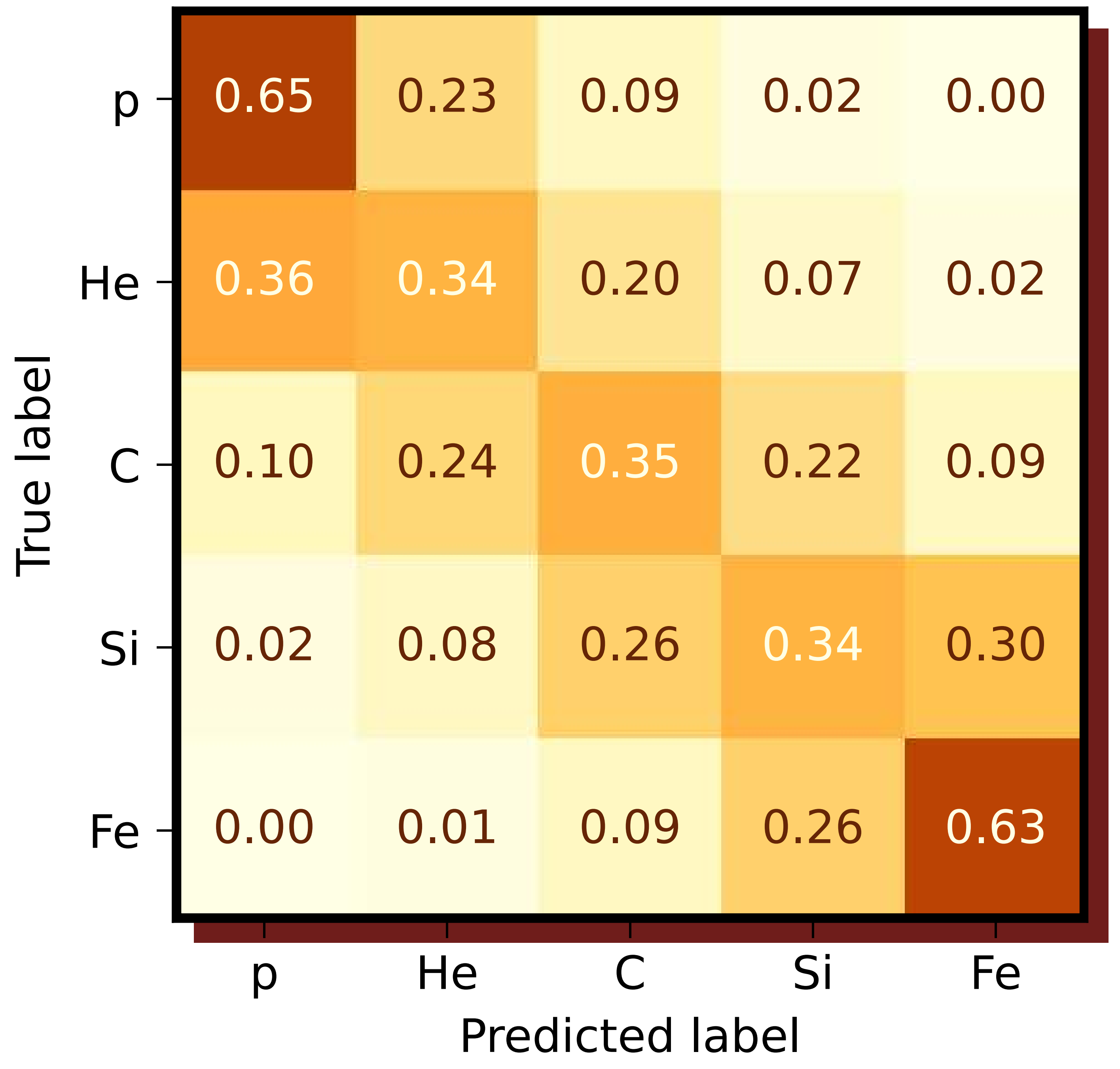
RF confusion matix



CNN confusion matix



MLP confusion matix



for QGSJet-II.04 hadronic interation model
(here and further extra cut at $\log_{10} E$, [eV] > 15.15)

Tests of the ML methods

Ablation study (impact of individual features on the result)

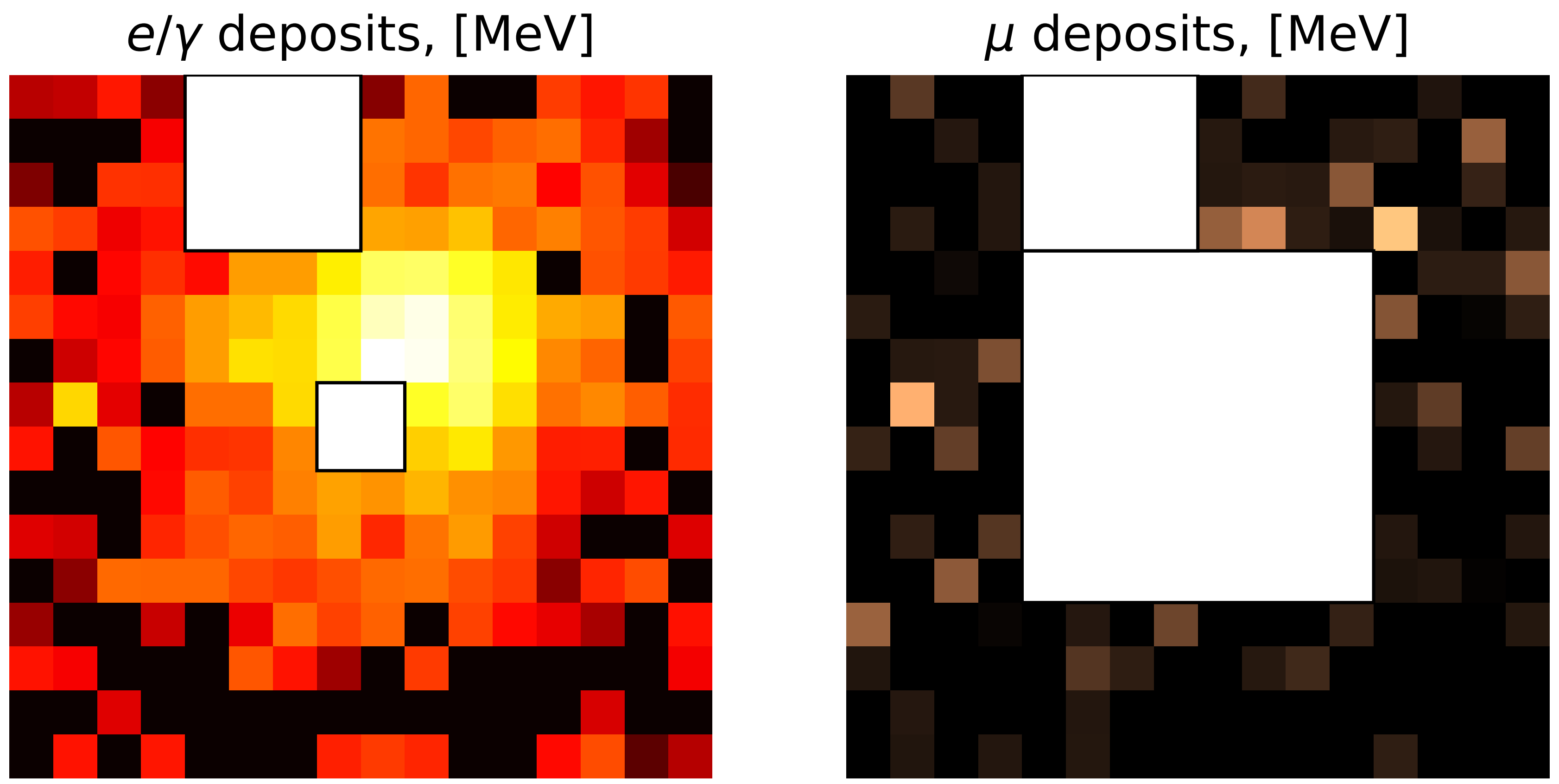
Train and test CNN with deposits only and reconstructed features only

CNN is stable with exclusion features except for the zenith angle.

Missing detectors study

Compare results of CNN on default dataset and “corrupted” dataset (with missing detectors)

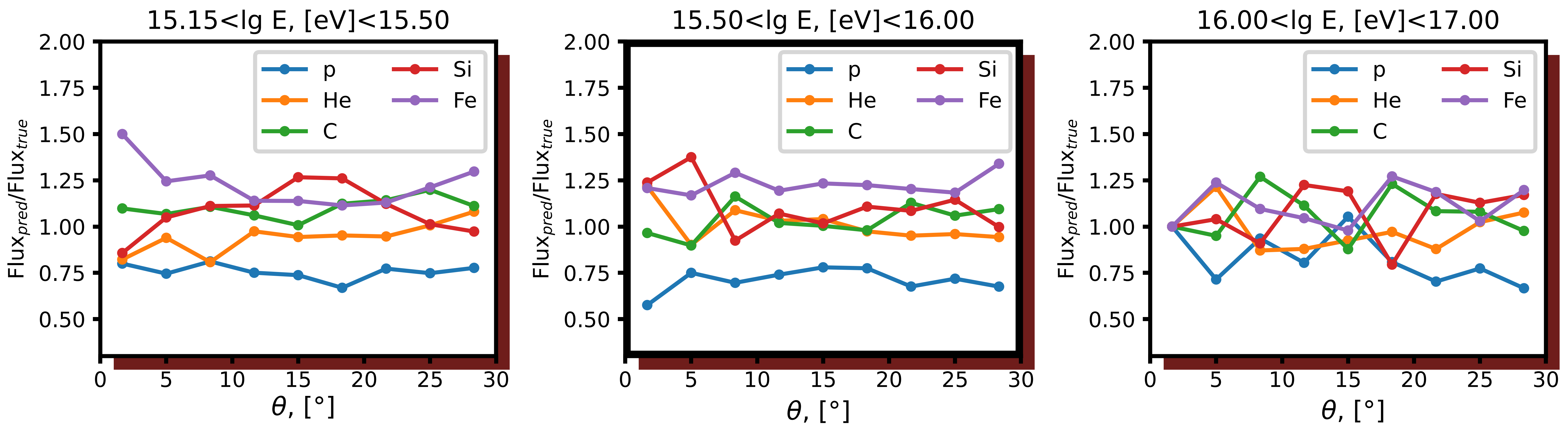
Decrease of diagonal elements of the confusion matrices by up to 4%



Example of the MC event for missing detectors study

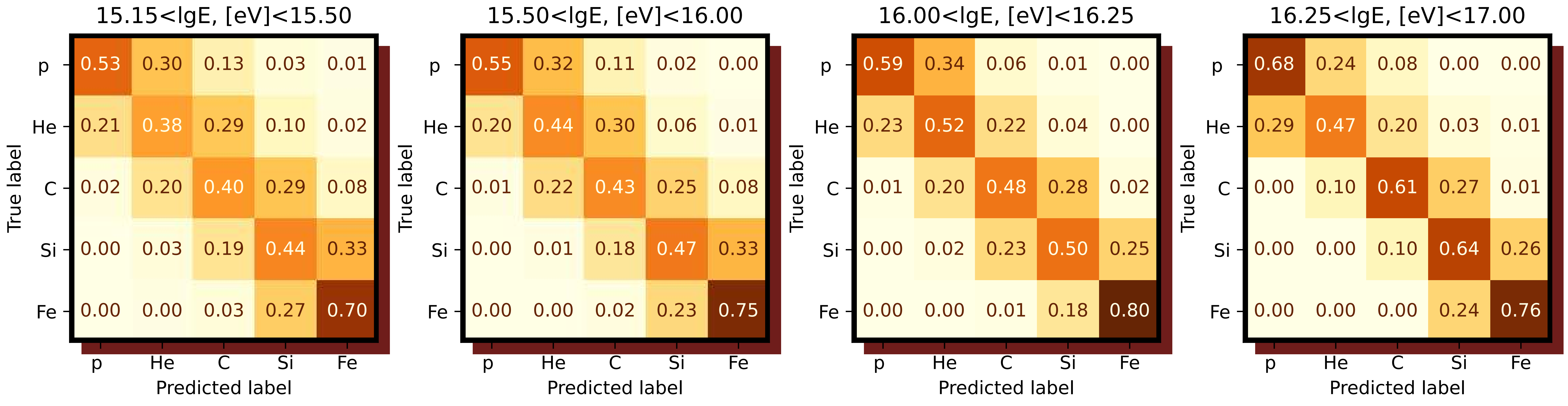
Tests of the ML methods

Zenith angle dependence



There is no clear dependence of the classifier prediction on the zenith angle

Energy dependence



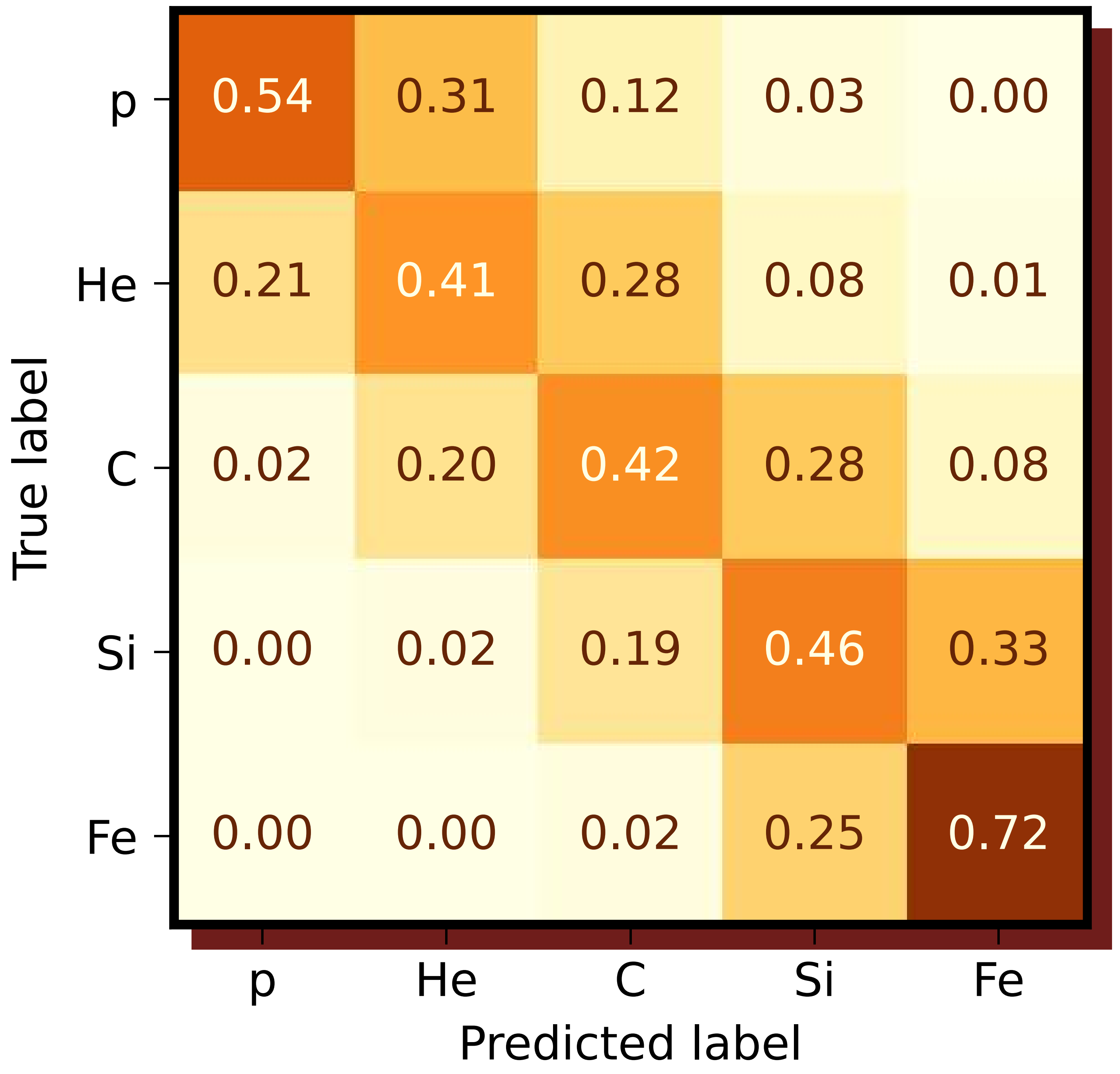
The more energetic showers are better classified

Tests of the ML methods

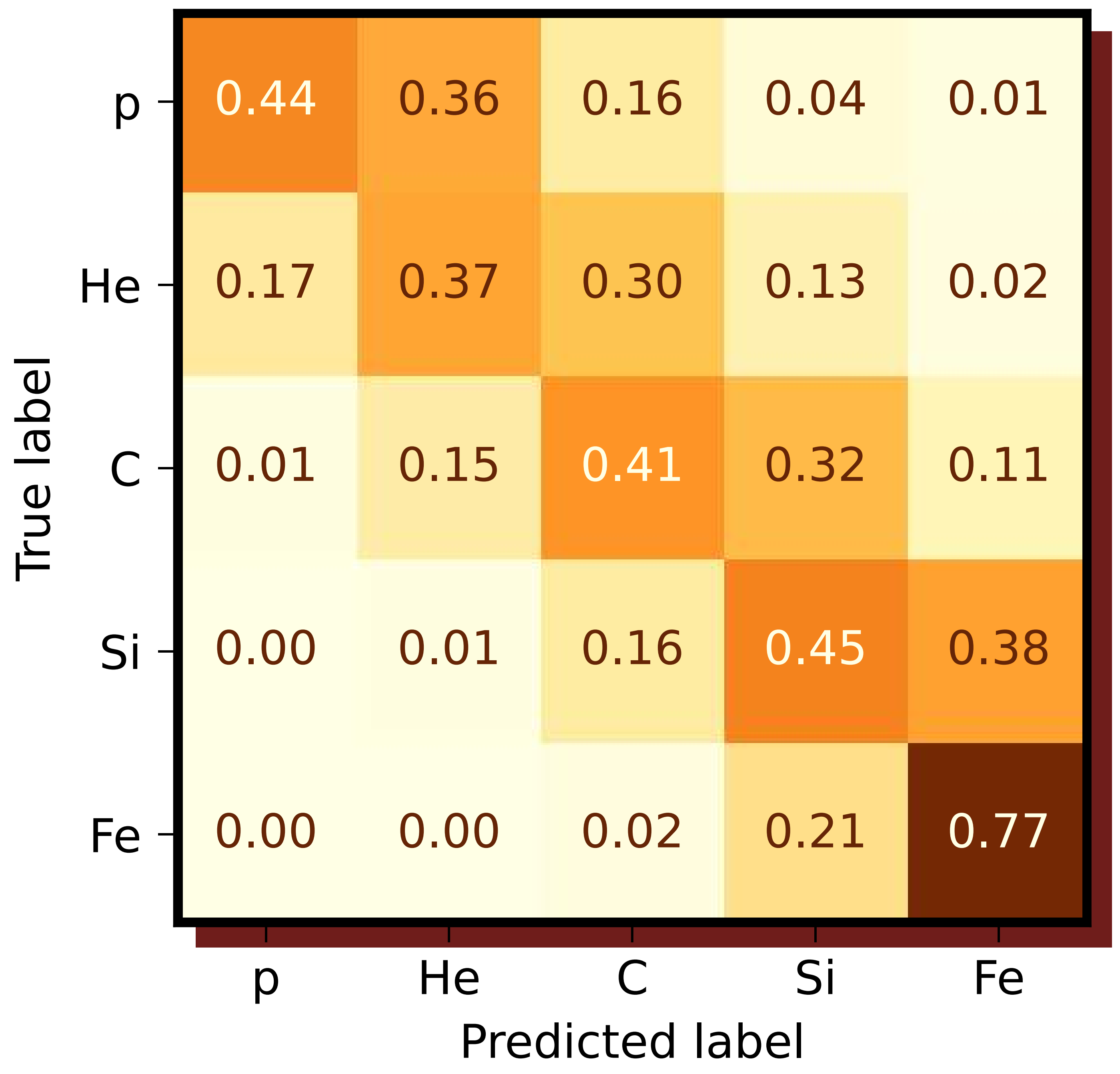
Cross-hadronic model reconstruction

Test the same CNN (trained on QGSJet-II.04) on different hadronic models

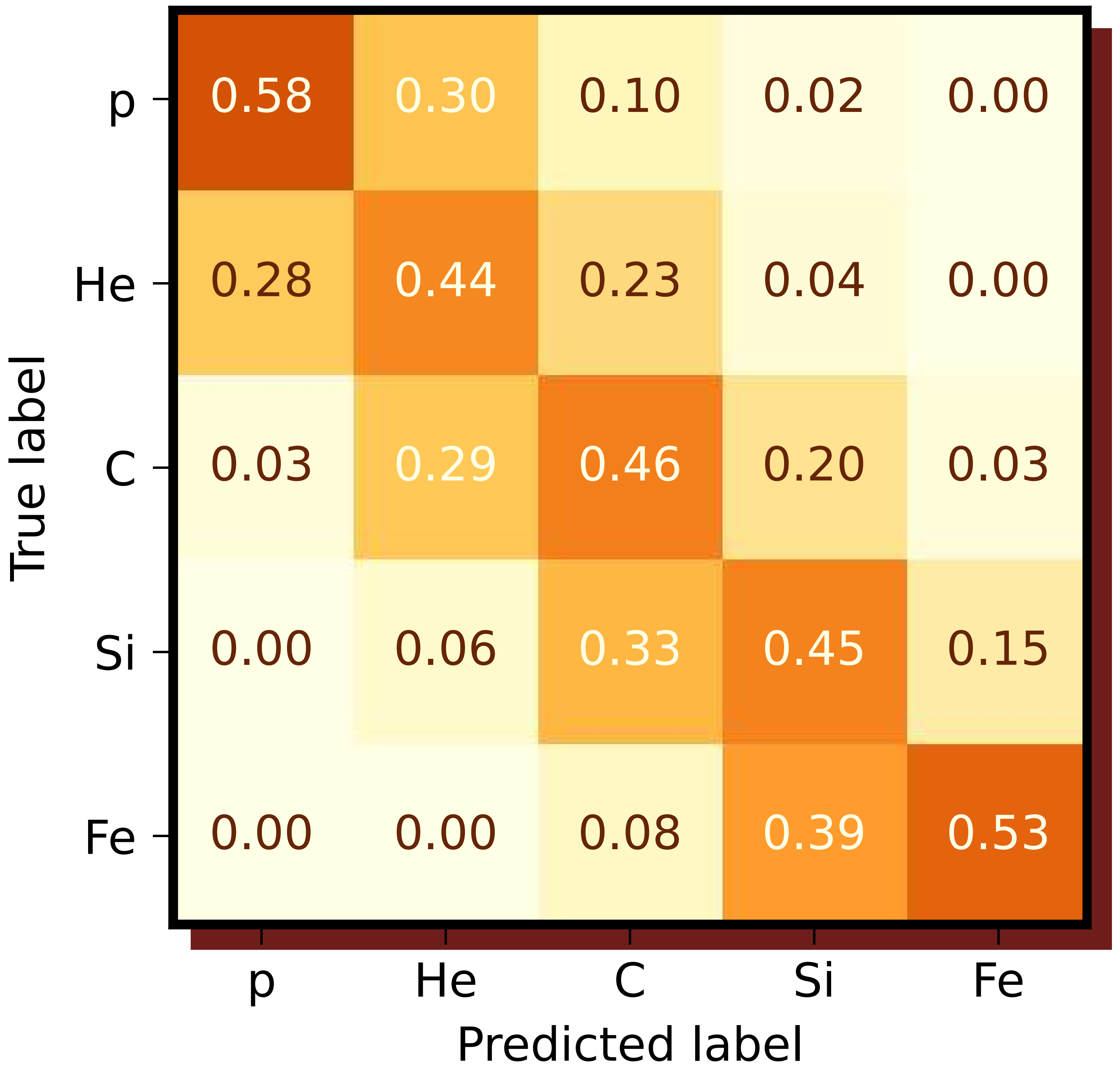
Test: QGSJet-II.04



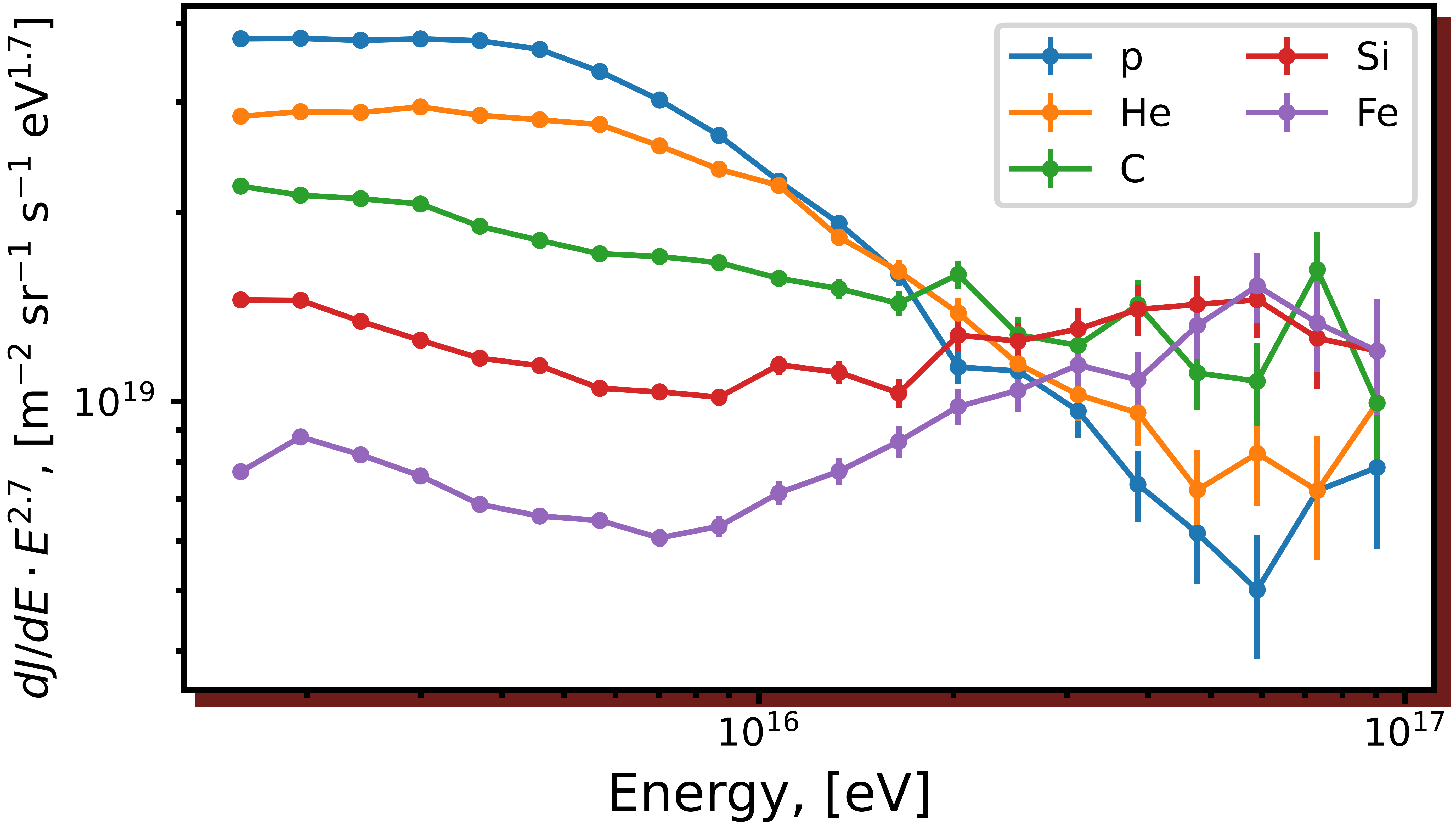
Test: EPOS-LHC



Test: Sybill 2.3c



Unblind folded result



Mass composition spectra (folded) on unblind data for the CNN trained with QGSJet-II.04

- Folded spectra means the spectra obtained by the direct predictions of the classifier
- Unblind set is 20% of the total experimental data

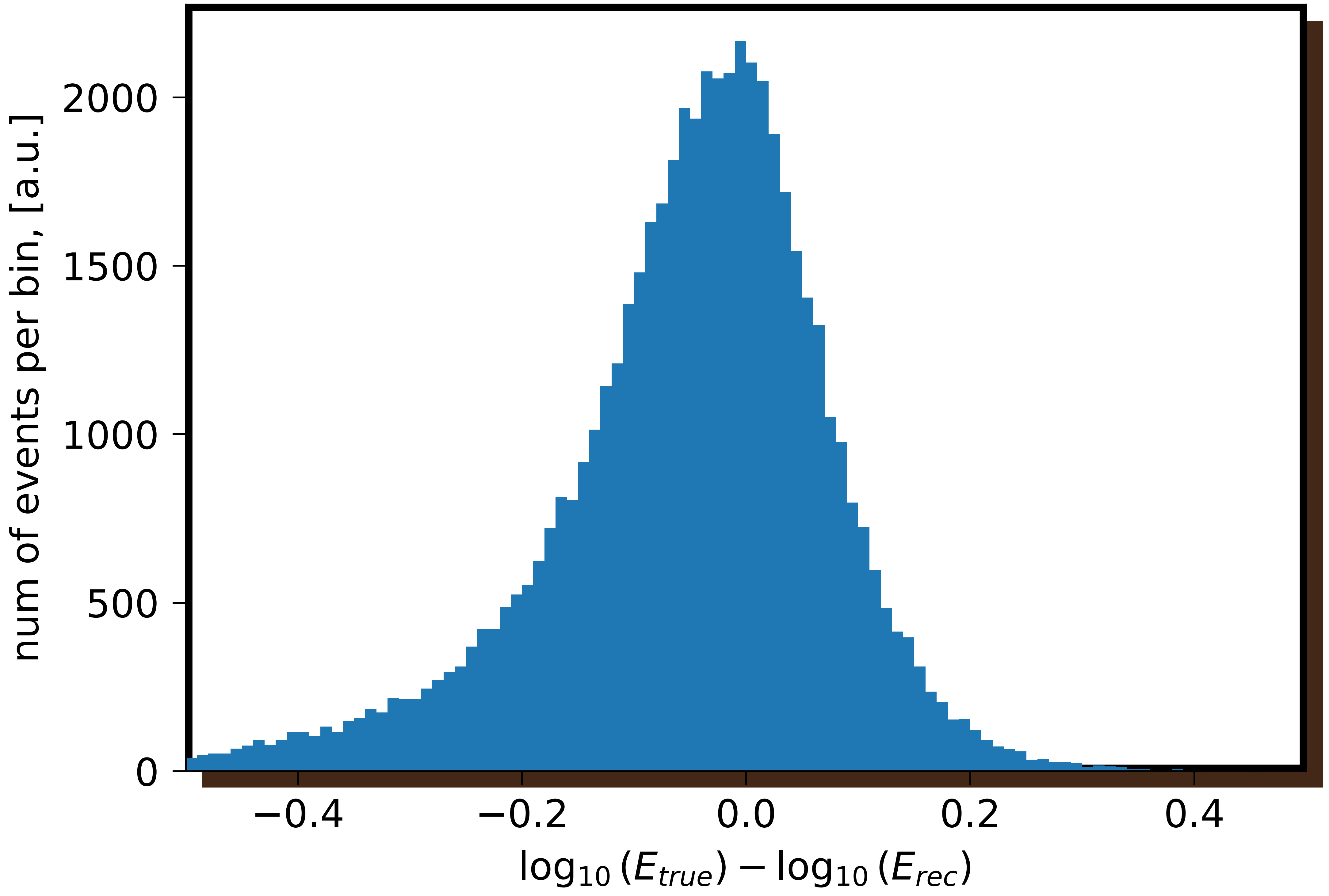
Unfolding procedure

Unfolding is a correction to the confusion matrix

We reconstruct mass composition spectra with unfolding procedure

We apply consequently two unfoldings:
- energy unfolding and
- particle type unfolding

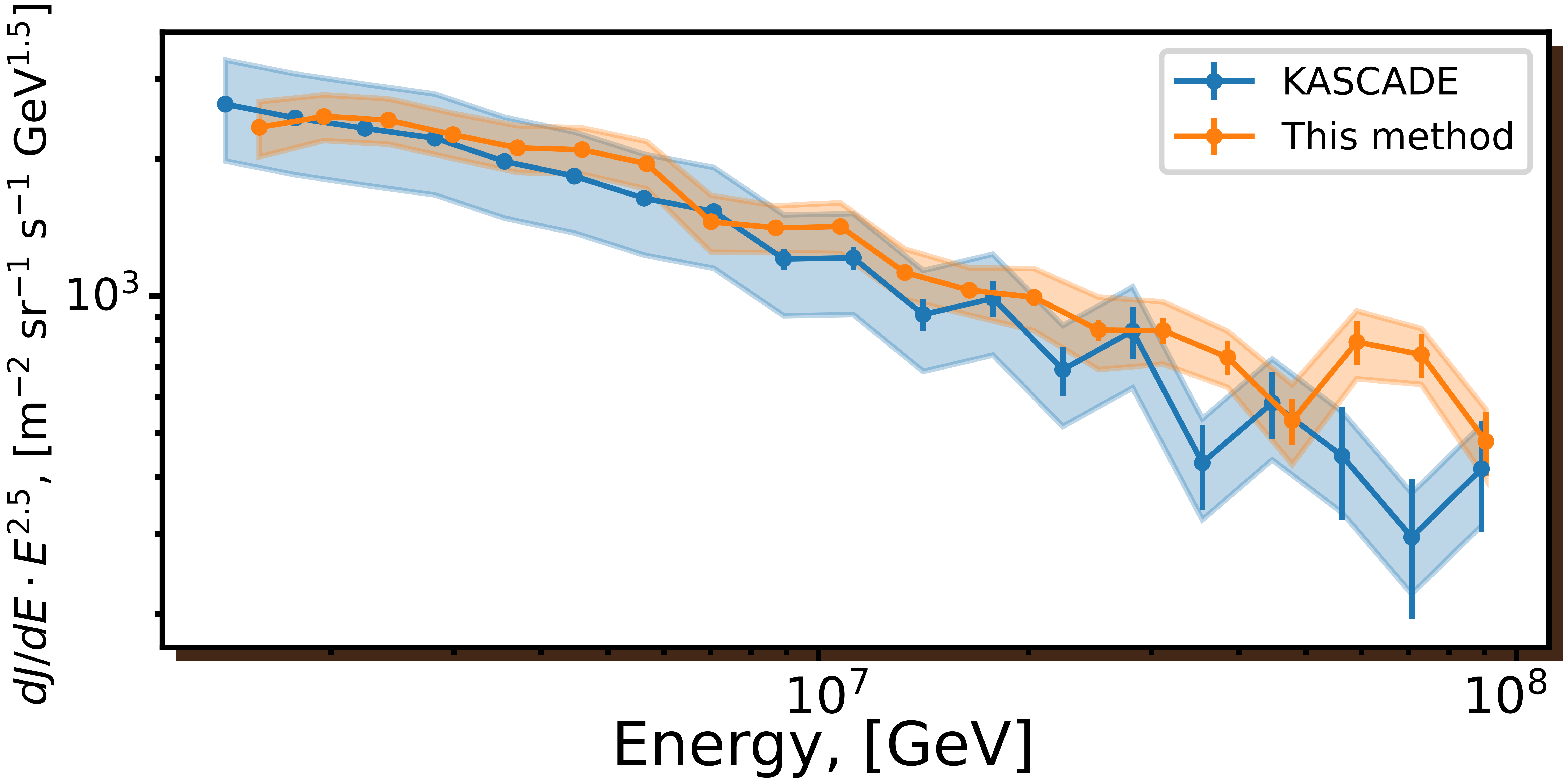
We use iterative bayesian unfolding method from pyunfold* package



Energy resolution of the KASCADE standard energy reconstruction for QGSJet-II.02 hadronic interaction model.

* Bourbeau et al., (2018). PyUnfold: A Python package for iterative unfolding. Journal of Open Source Software, 3(26), 741, https://doi.org/10.21105/joss.00741

Total flux comparison



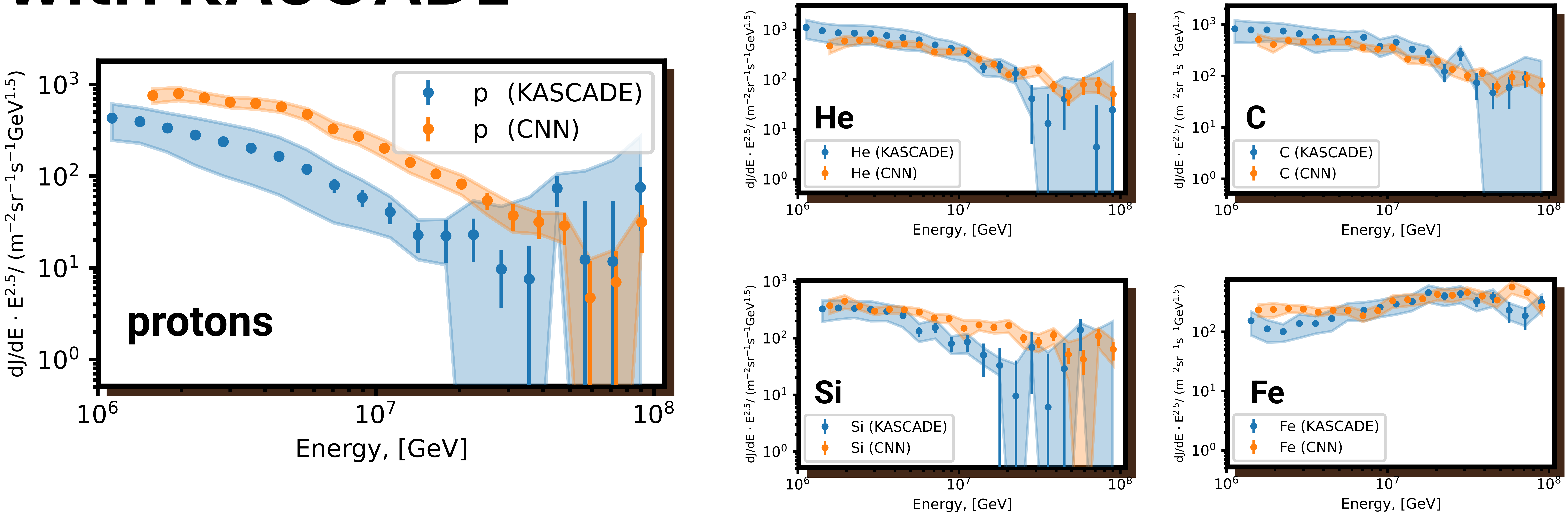
Total spectra comparison with energy unfolding for our method and original KASCADE results* with QGSJet-II.02 hadronic interaction model (unblind data).

Whiskers are statistical uncertainties

Bands include systematic uncertanties of:
- limited MC,
- missing detectors,
- unknown composition prior,
- spectra slope prior,
- the unfolding method

* W. D. Apel et al., KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays, Astropart. Phys. 47 (2013) 54−66, [1306.6283]

Mass components spectra comparison with KASCADE

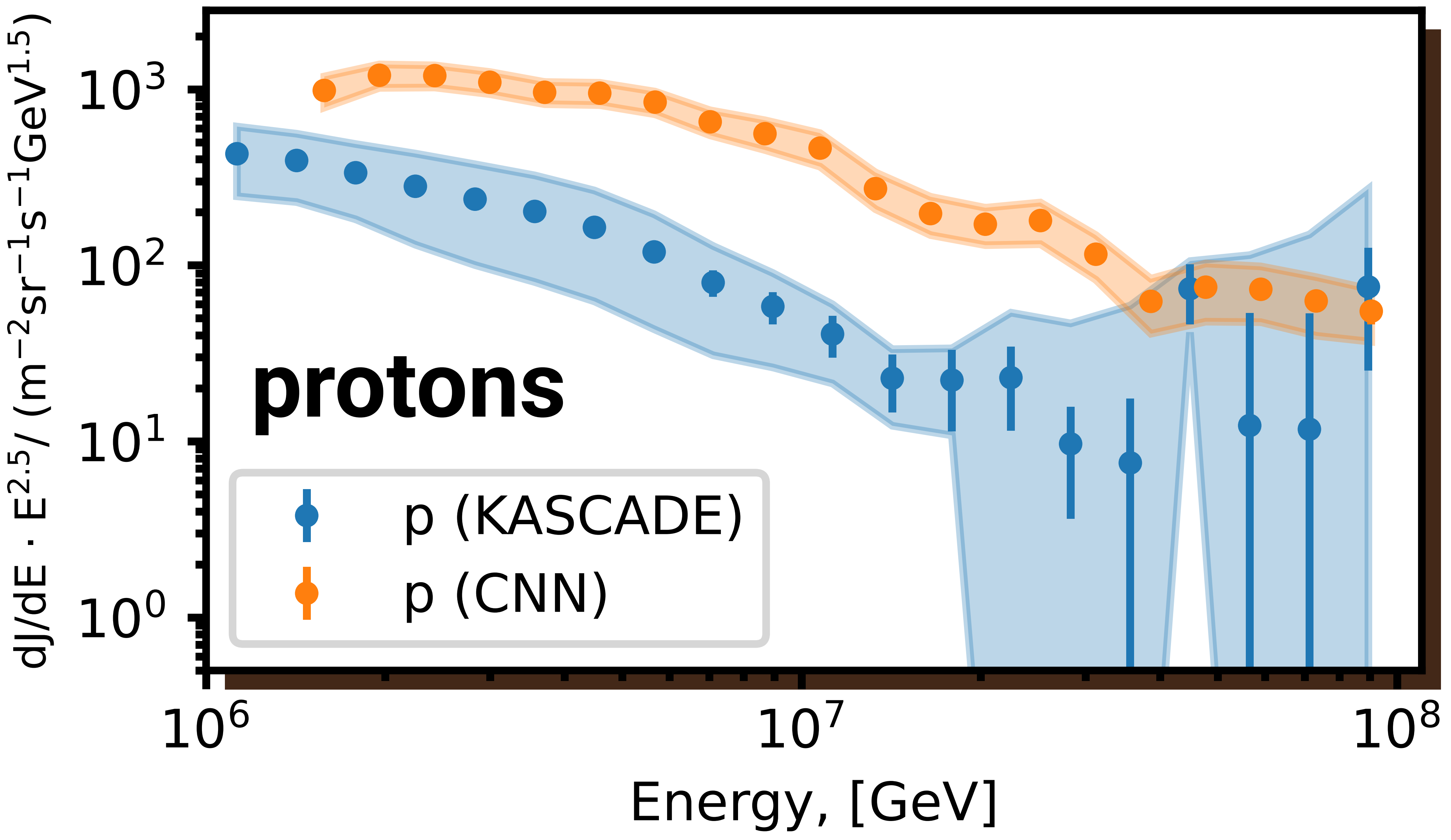


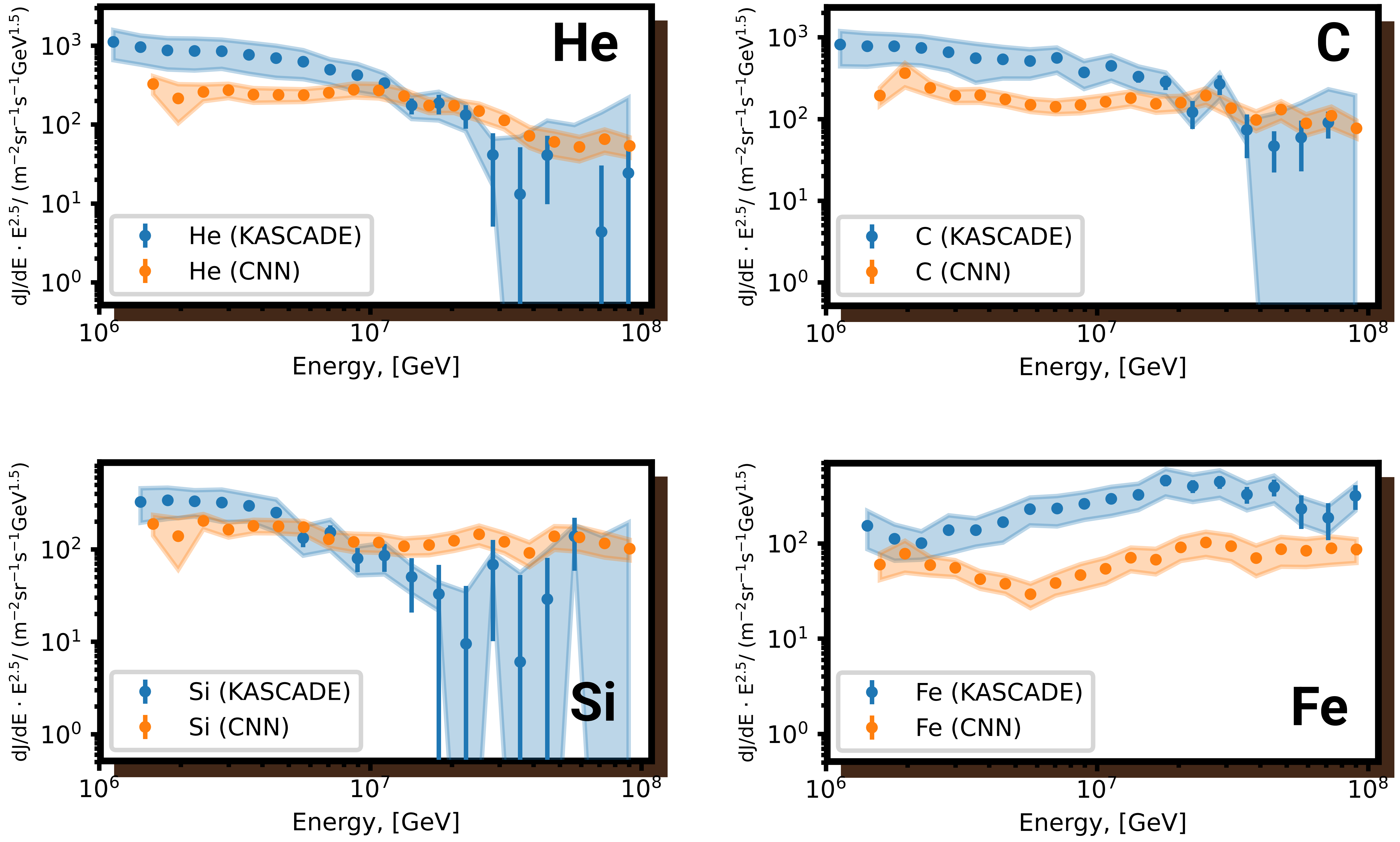
One-to-one comparison of the our spectra (orange, unblind data) and original KASCADE results* (blue, QGSJet-II.02 hadronic interaction model)

* Apel, W. D. et al. (2013). KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays. Astroparticle Physics, 47, 54–66. doi:10.1016/j.astropartphys.2013.06.004

Results. QGSJet-II.04

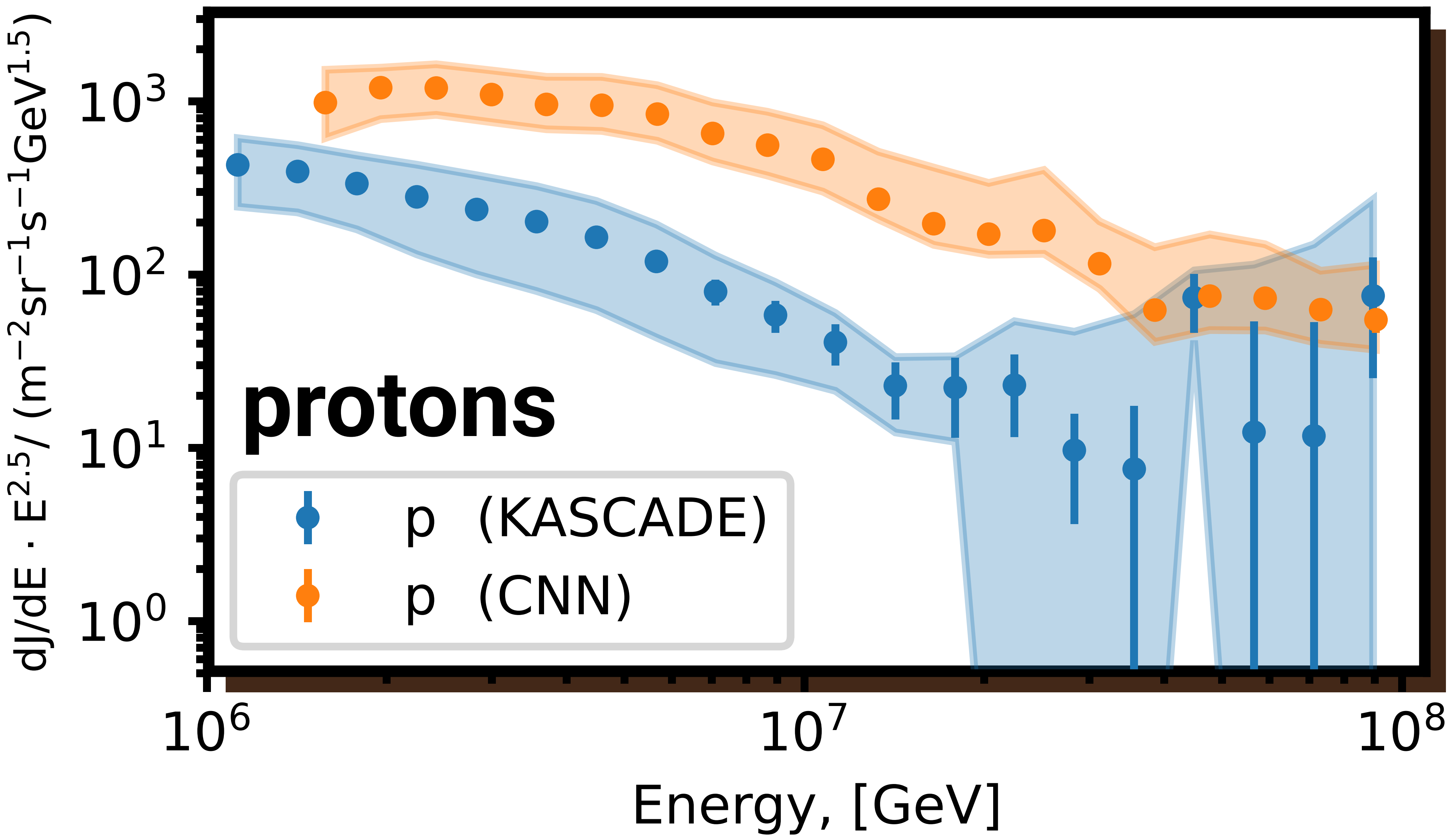
Orange: reconstructed spectra for QGSJet-II.04 hadronic interaction model on blind data

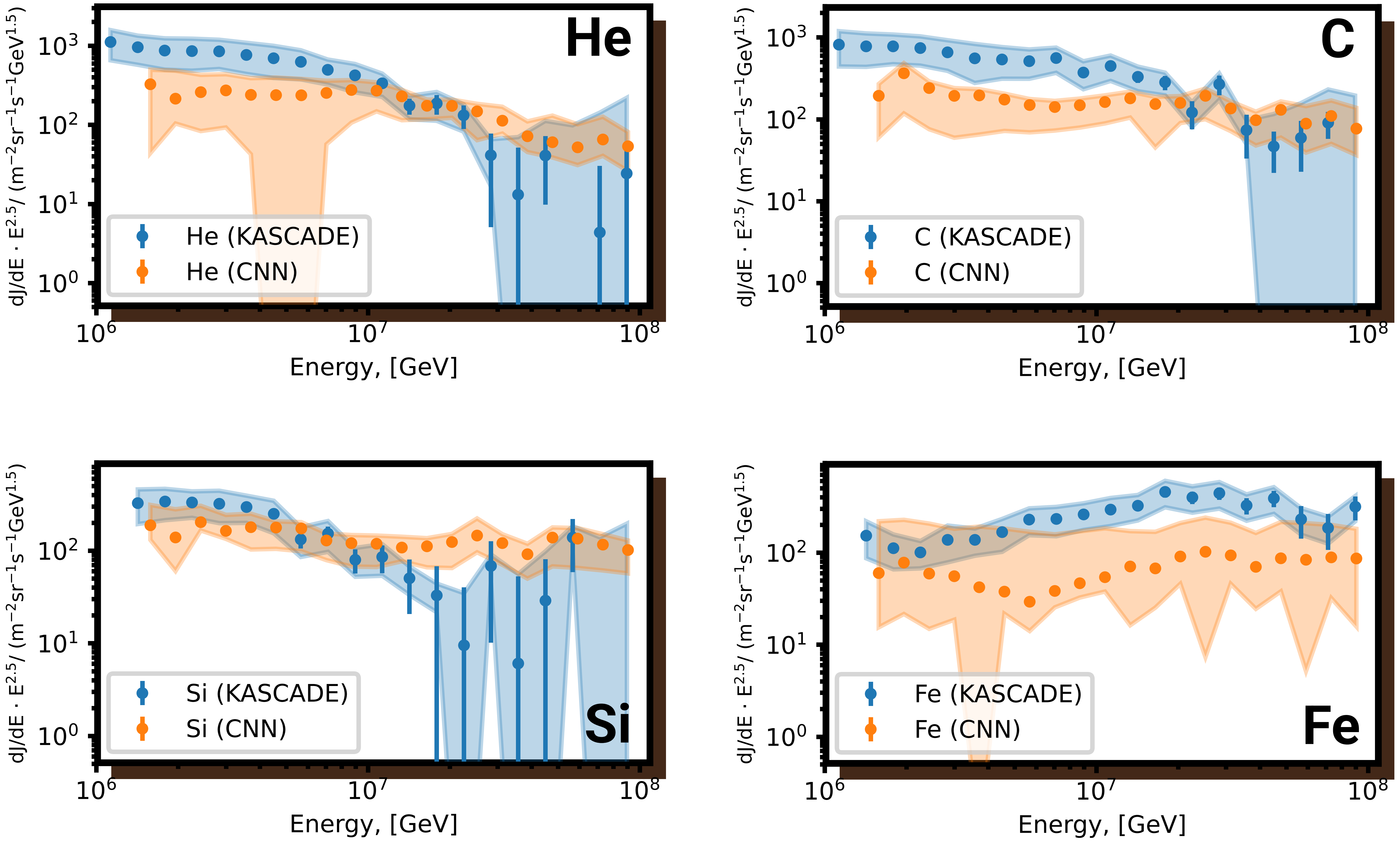




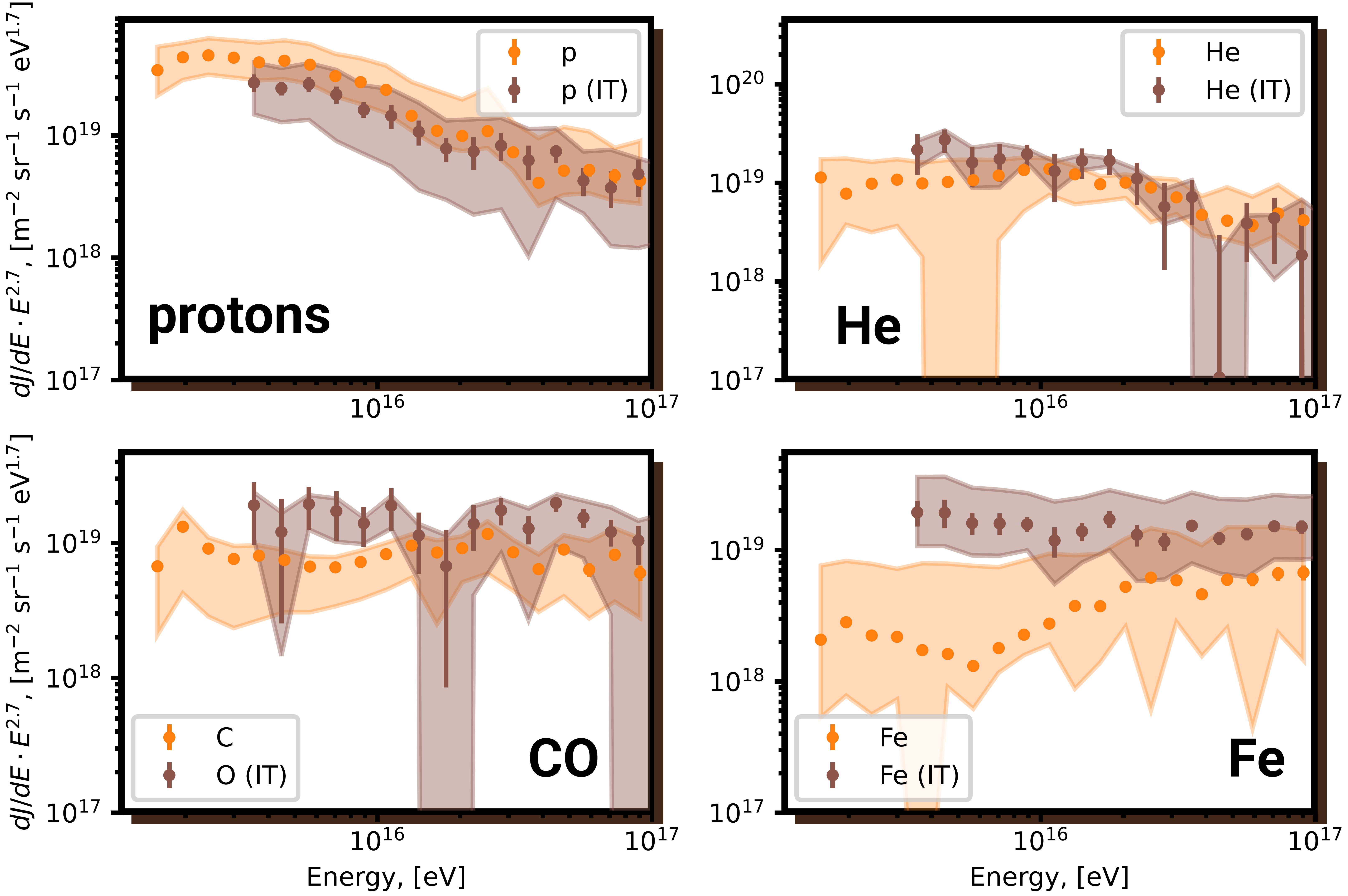
Results. Cross-hadronic systematics

Orange: reconstructed spectra for QGSJet-II.04 hadronic interaction model on blind data with cross-hadronic model systematics





Results. IceTop comparison



Orange: reconstructed spectra for QGSJet-II.04 hadronic interaction model on blind data with cross-hadronic model systematics

Brown: IceTop results* (Sybill 2.1)

* Aartsen, M., & others (2019). Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube. Phys. Rev. D, 100(8), 082002.

Conclusion

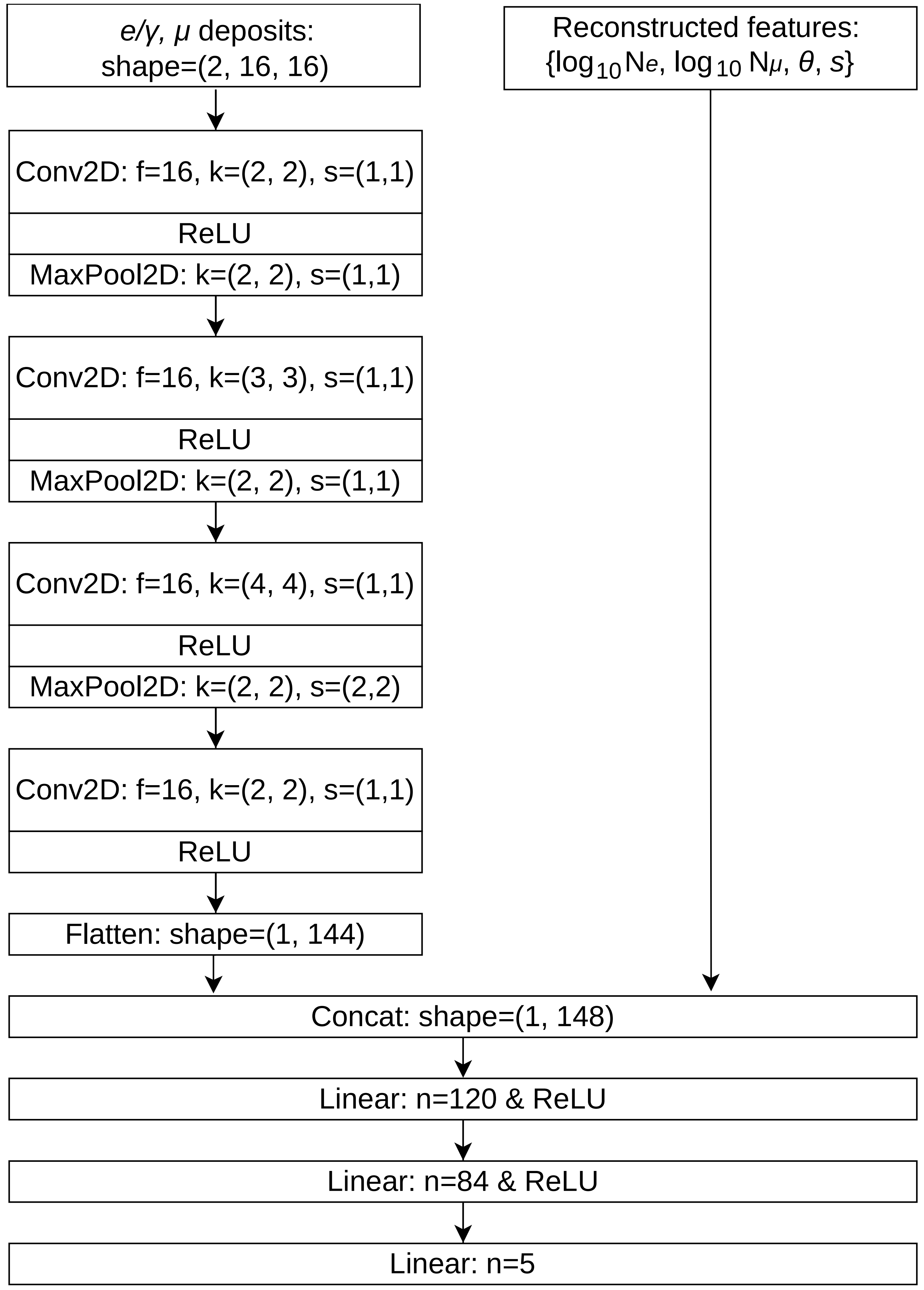
- We reanalyzed data of KASCADE cosmic ray experiment
- We reconstructed cosmic ray mass components spectra for new hadronic interaction models and took into account cross-hadronic model systematics
- General uncertainties of the our method are much smaller than those of the standard KASCADE reconstruction
- We found a significant dominance of the proton component
- We have a general agreement with IceTop

Thanks for attention!

Supported by the Russian Science Foundation (RSF), grant 22-22-00883

Backup slides

Architectures



CNN

MLP