ER/NR Discrimination

Content

- Digitisation and Reconstruction of MC
- Linearity and Resolution
- ER/NR Discrimination using ML
- Application to LIME data
- Drift Distance Estimation using ML

ER

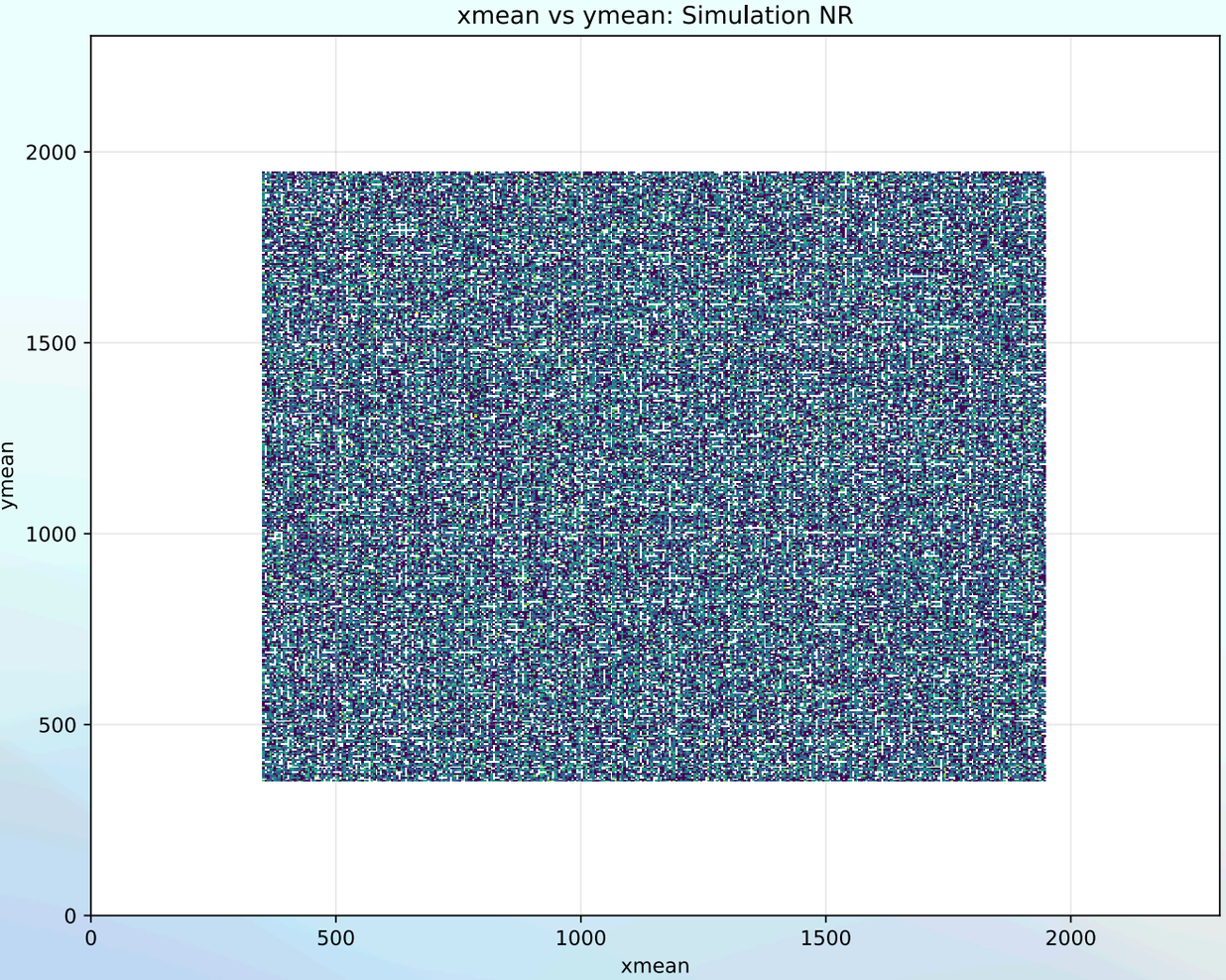
NR

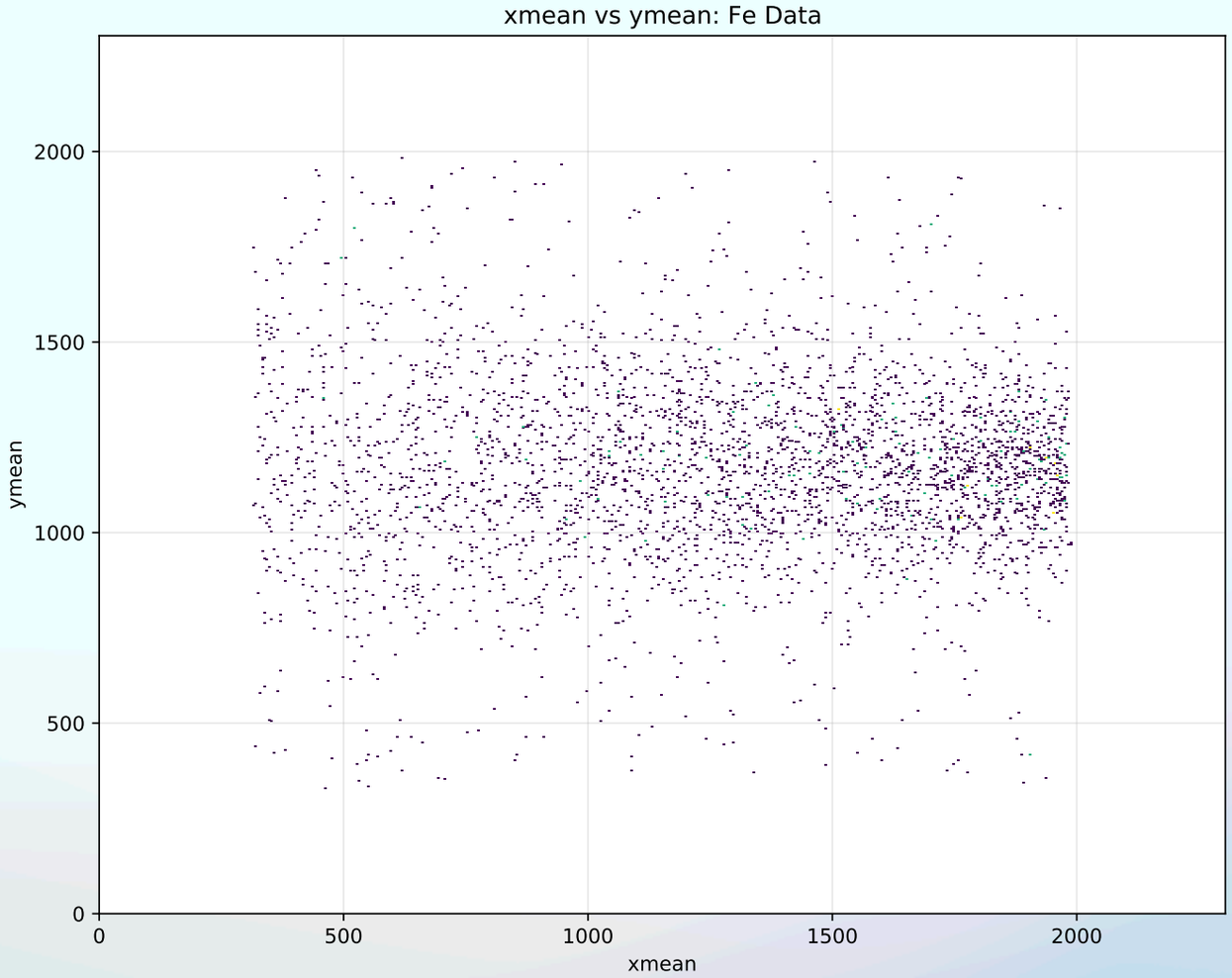
Simulation, Digitisation & Reconstruction

- Primary Simulation
  - For ER samples: Geant4
  - For NR simulation: SRIM
  - Energies: [2-50] keV in steps of 2 keV
  - # of samples: 10,000 at each energy
- Digitisation
  - All the samples are digitised at 800 V/cm drift field to match the LNGS LIME data.
- Reconstruction
  - Winter23 branch

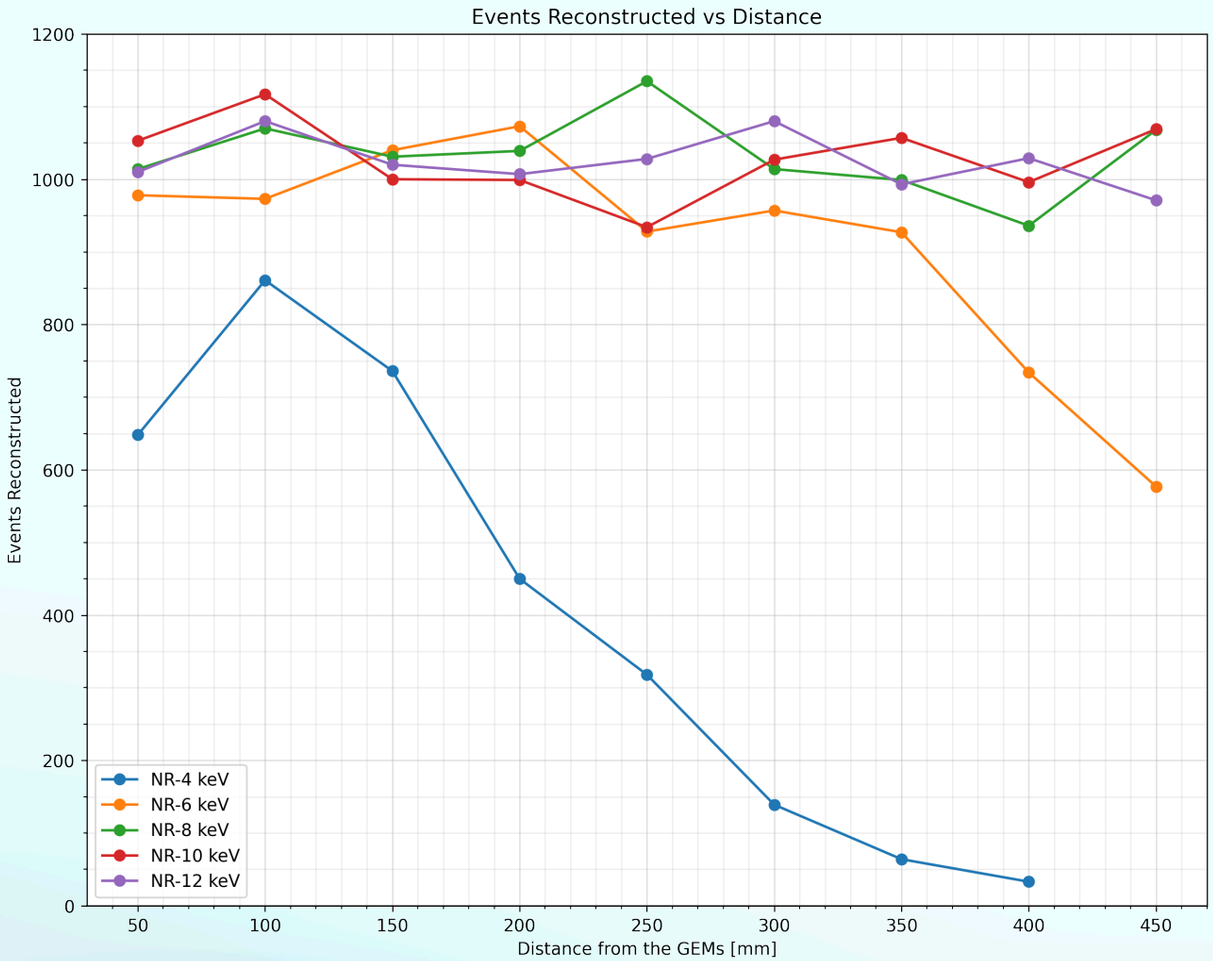
Event Selection

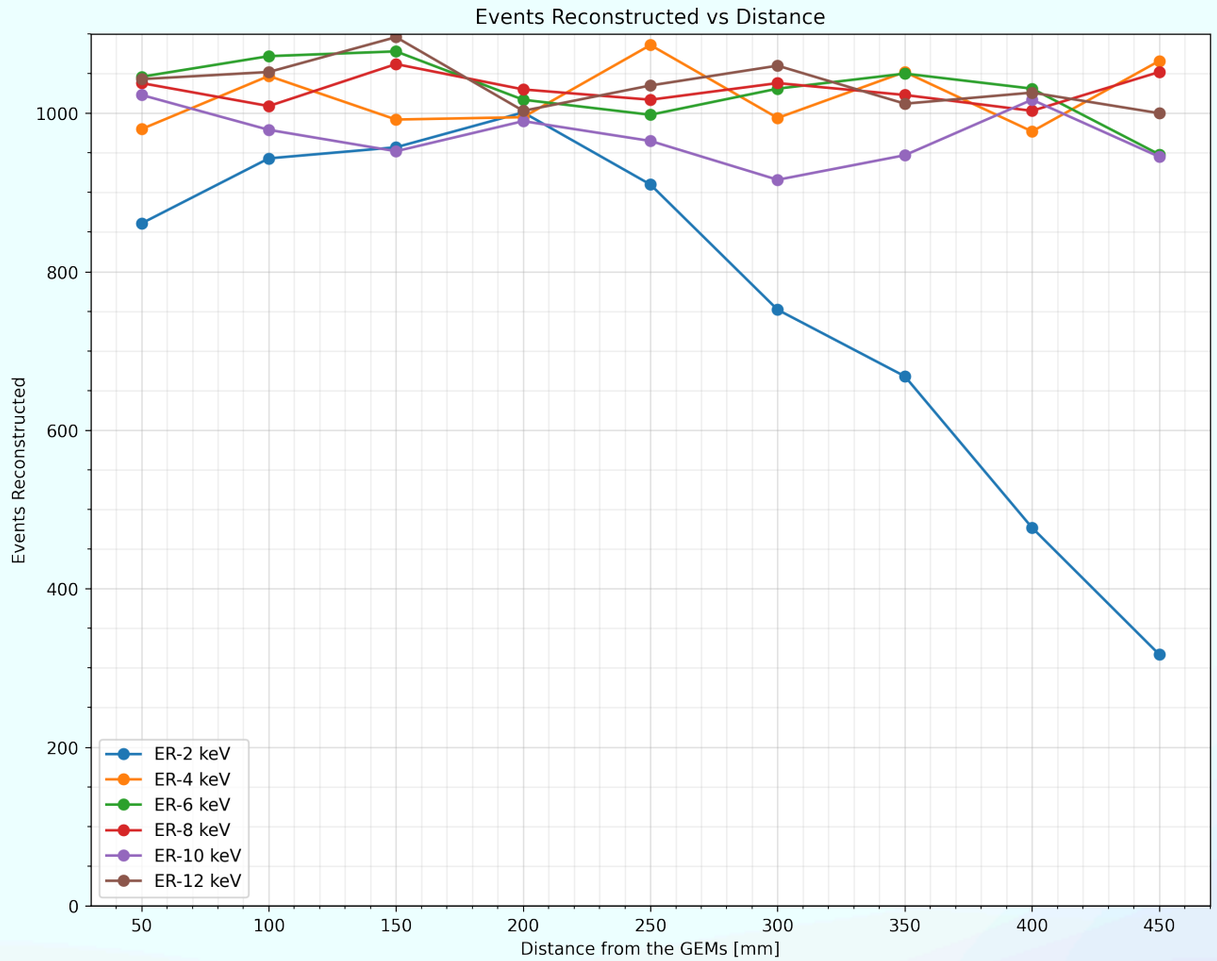
- Noise
  - rms > 6
  - tgausssigma * 0.152 > 0.5





Reconstructed events:



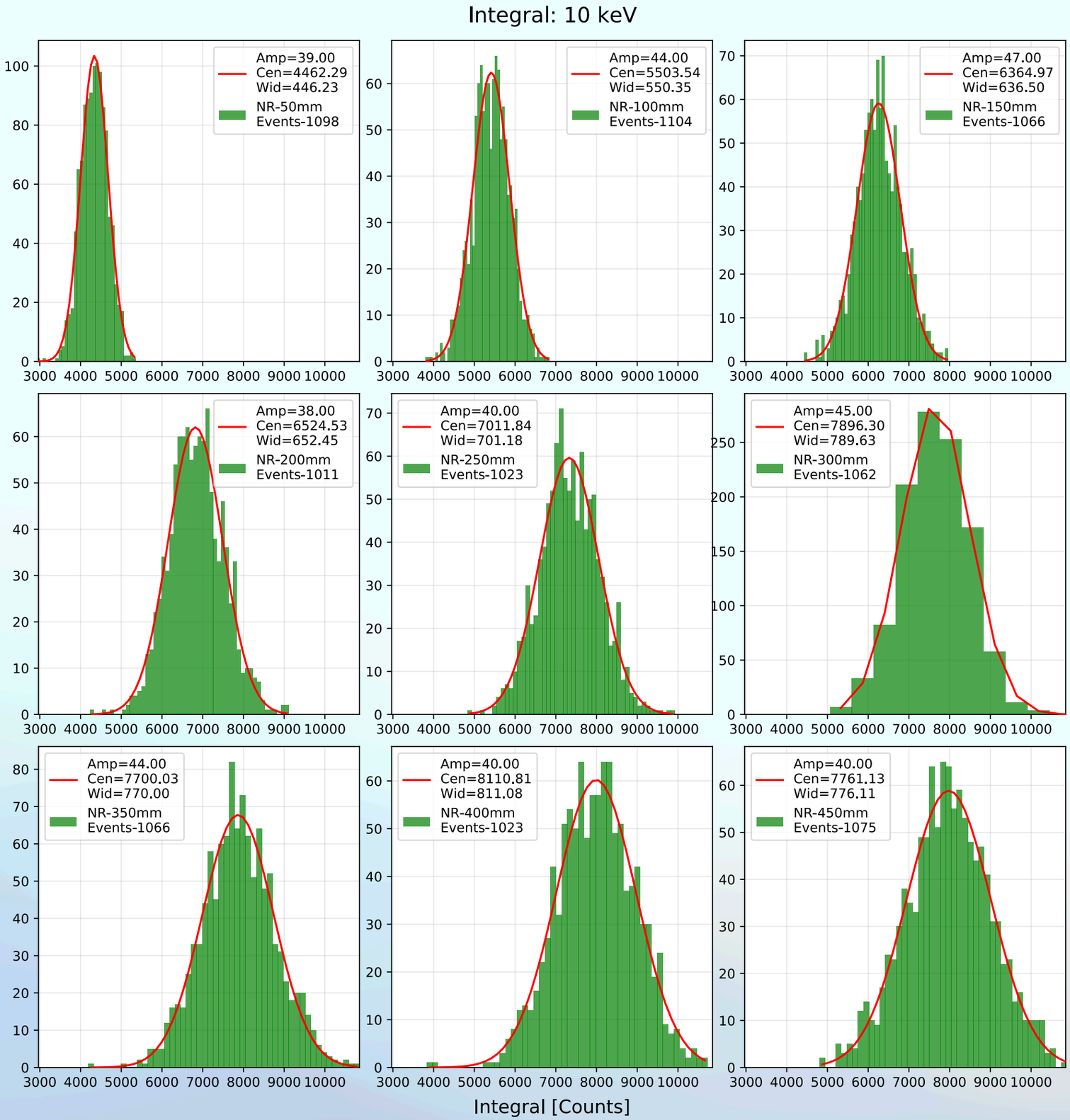


Energy_ER = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50]

Energy_NR = [4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50]

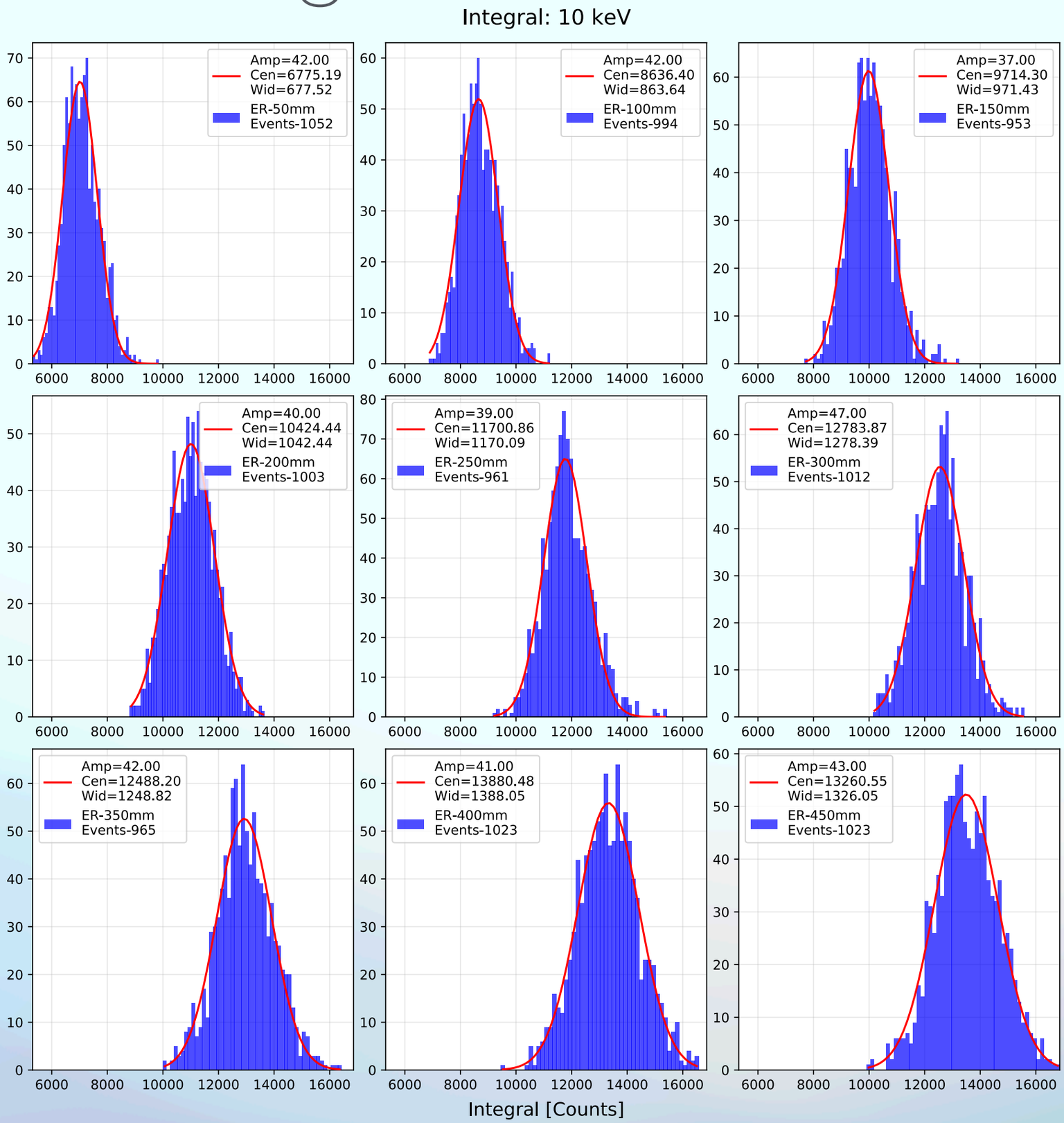
Energy_Quenched = [1.3, 2.5, 3.9 , 5.4, 7.0 ,8.7, 10.5, 12.3, 14.0, 15.9, 17.7, 19.6, 21.5, 23.4, 25.3 ,27.2, 29.2, 31.1, 33.0, 34.9 , 36.9, 38.9, 40.8, 42.8]

Integral fitting at different distance - NR

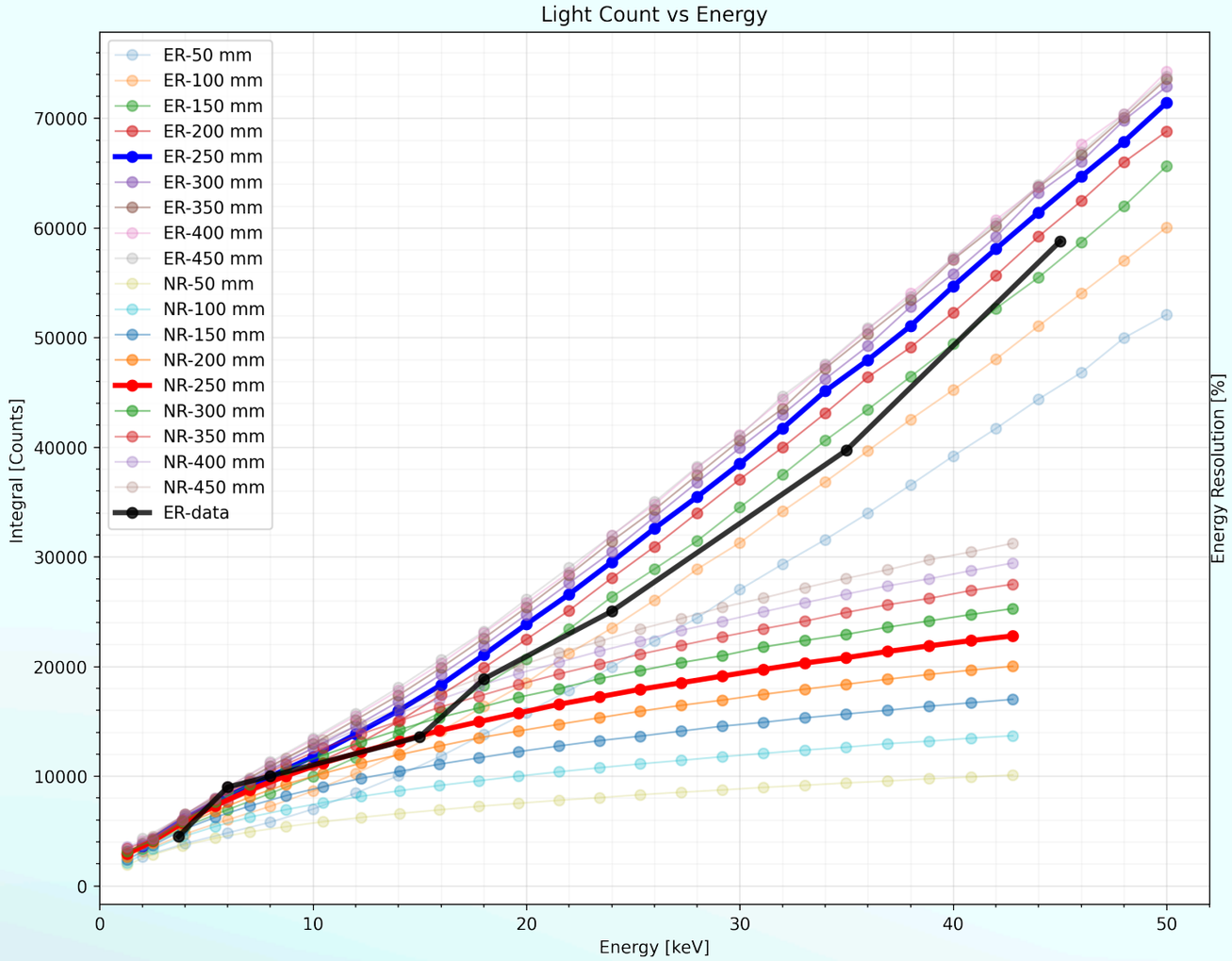


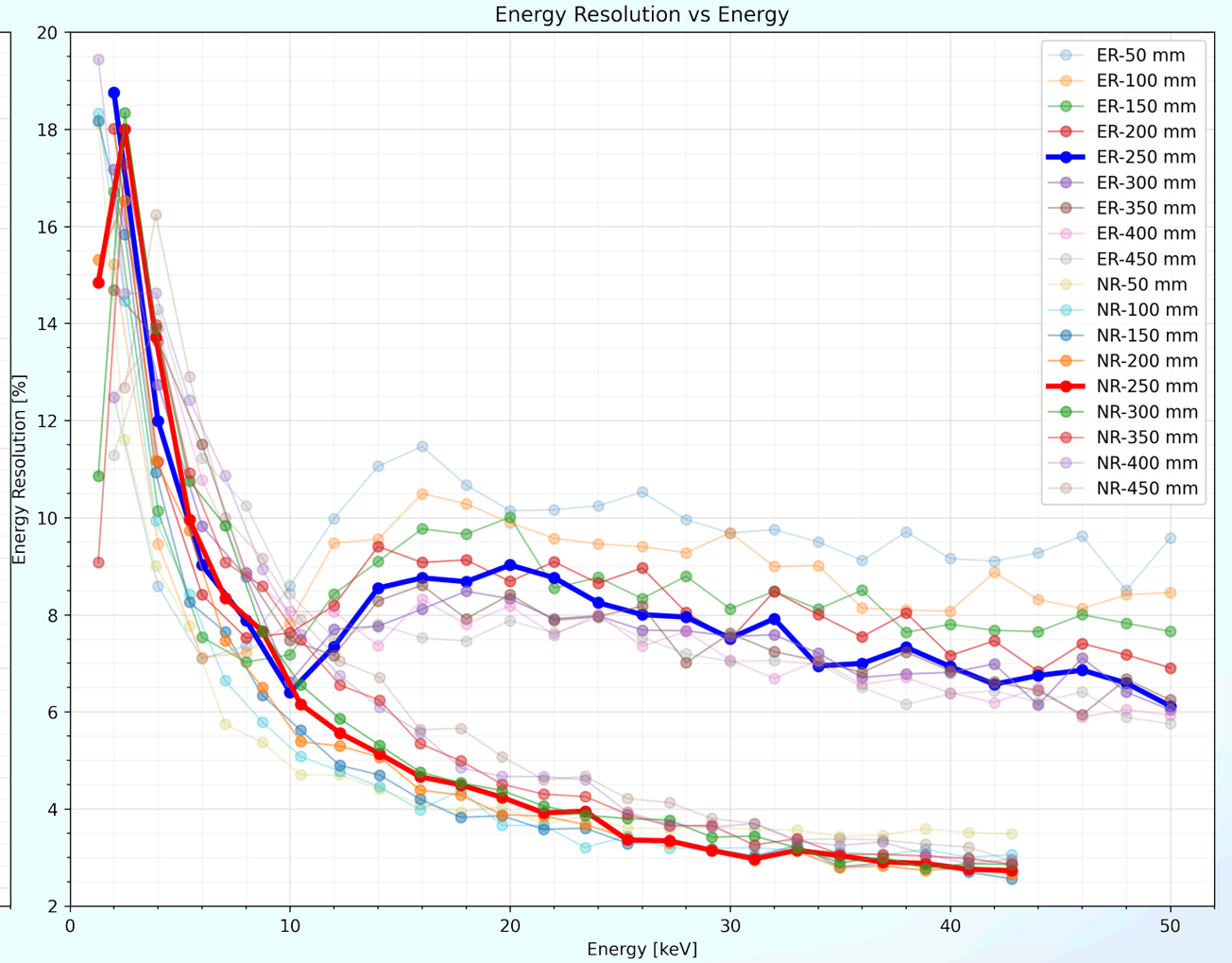
- NR integral distribution at different distances from the GEM fitted with Gaussian.
- As we go away from the GEMs integral increases indicating presence of saturation.

Integral fitting at different distance - ER



Energy Linearity & Resolution

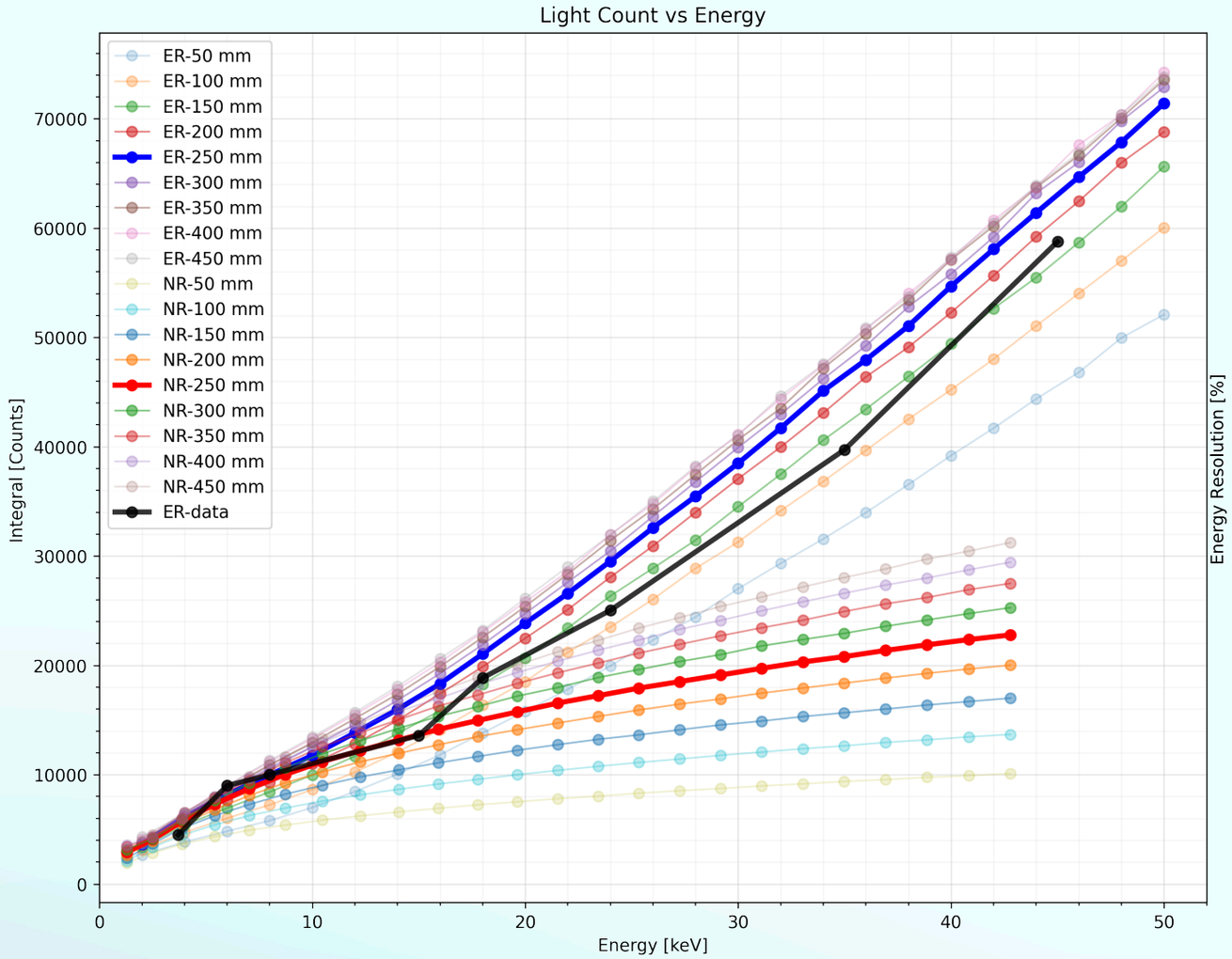


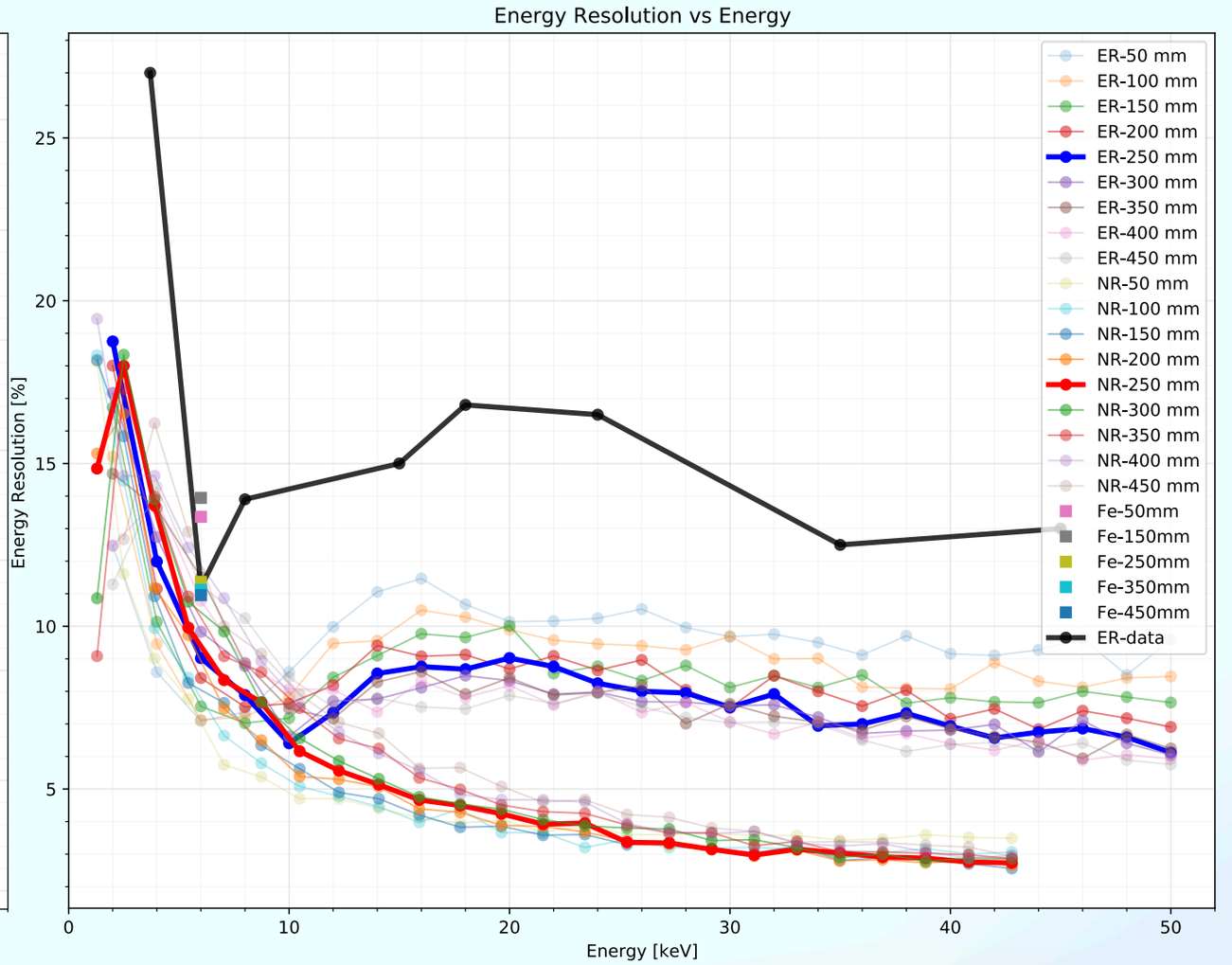


- ER shows good linearity and the effect of saturation is very low at distances greater then 25 cm.
- NR shows very high saturation and they are saturated almost at all energies and even at the farthest distance from the GEM.

* ER-data in the plot is Data taken with LIME at LNF.

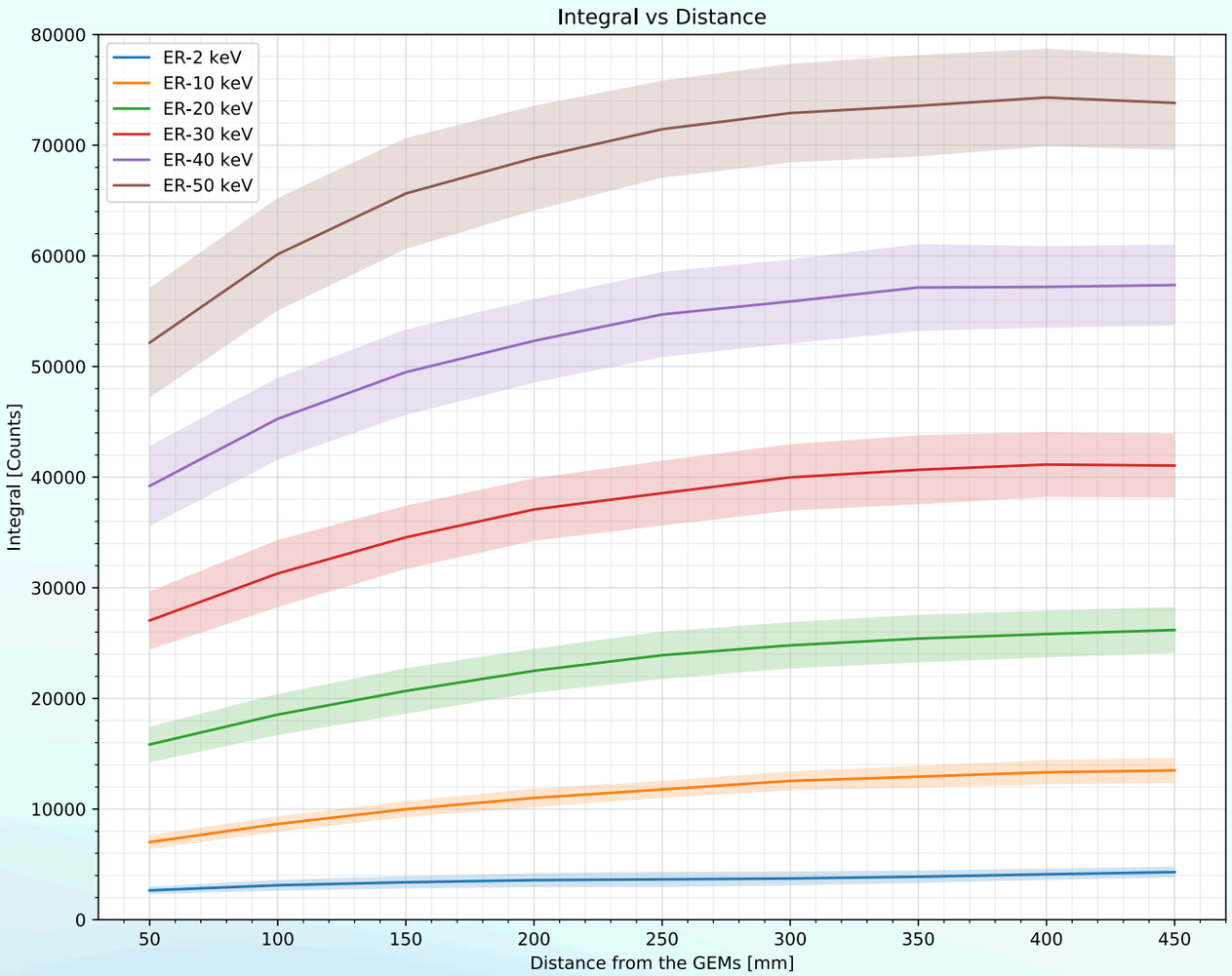
Energy Linearity & Resolution

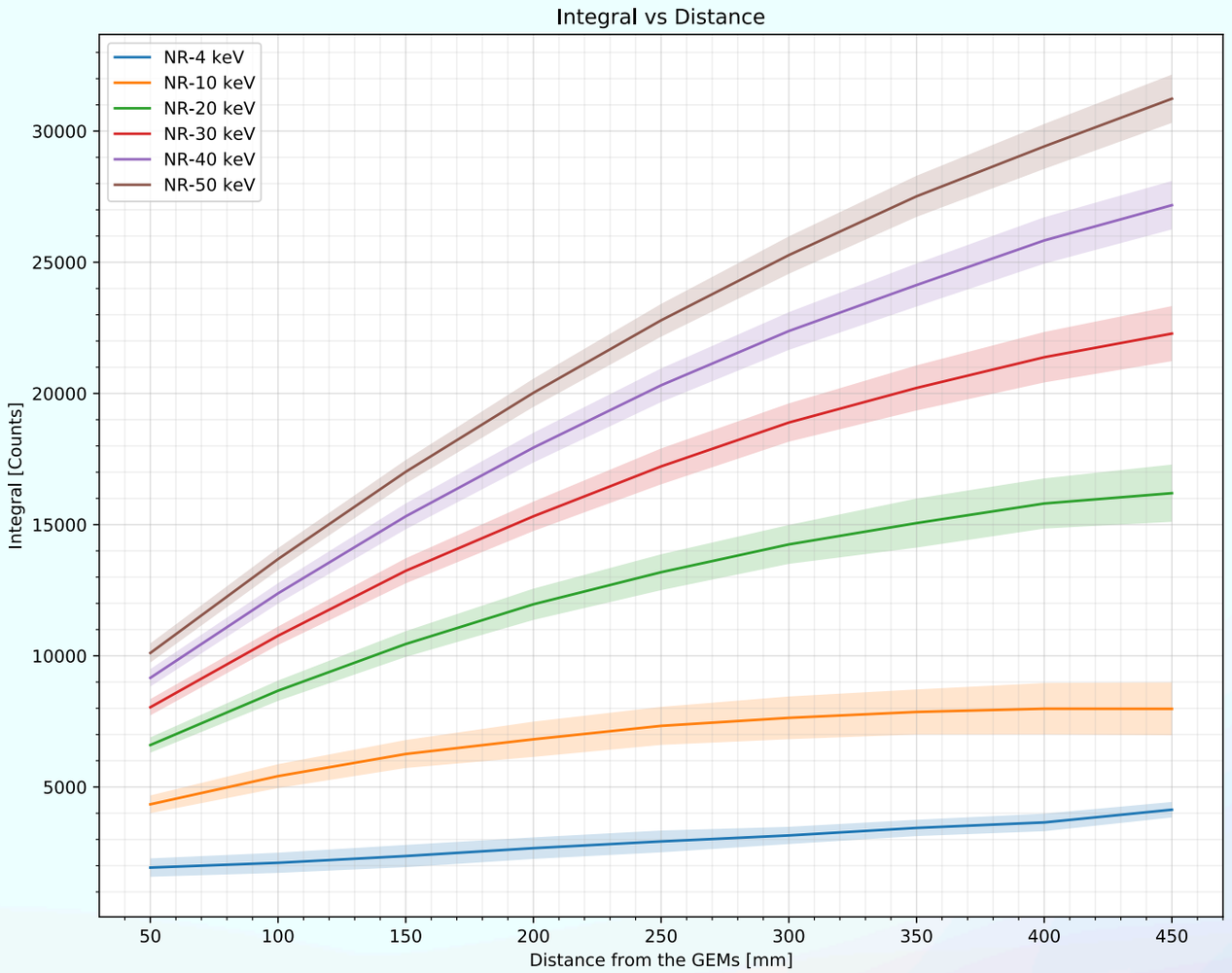




* ER-data in the plot is Data taken with LIME at LNF.

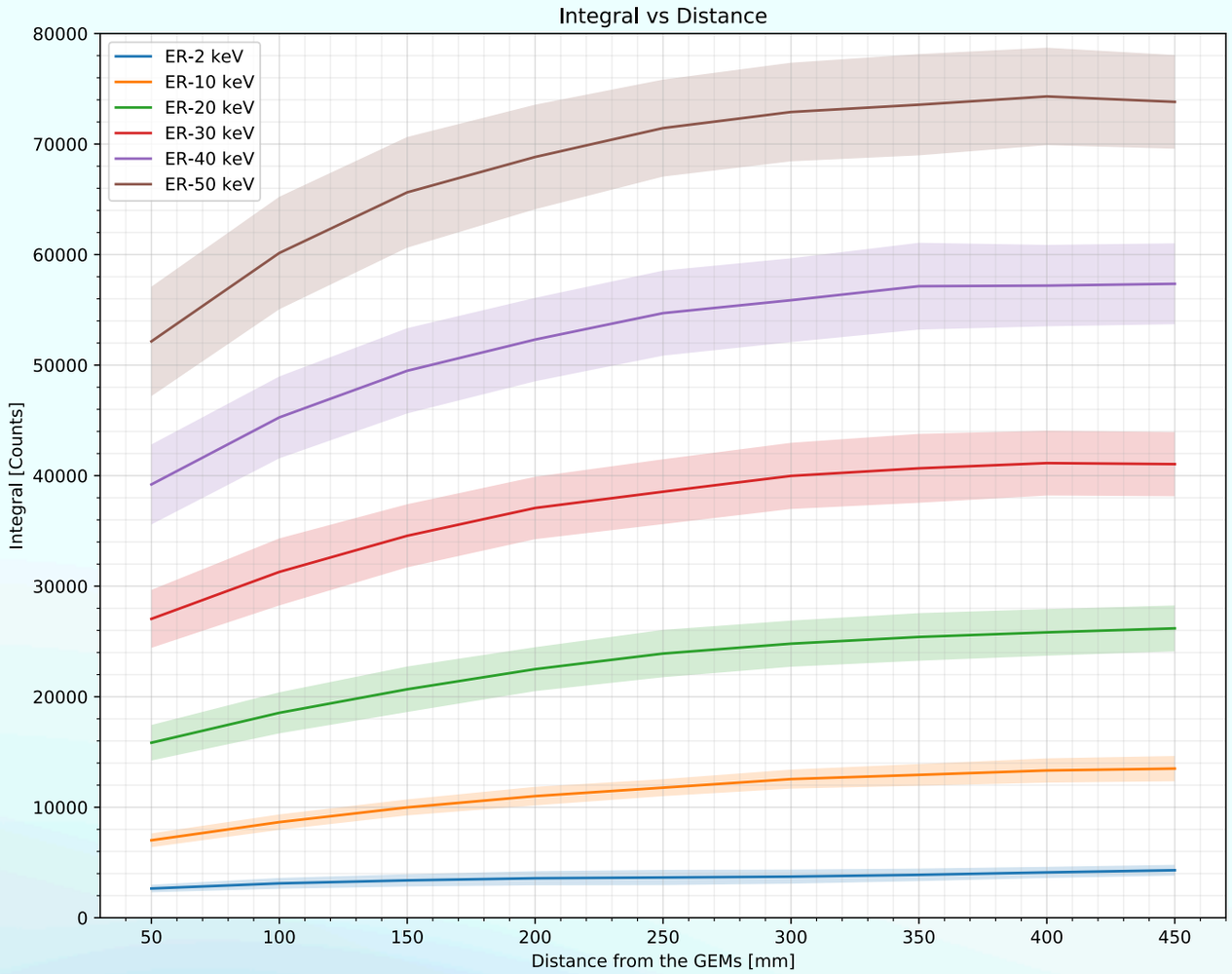
Integral vs Distance

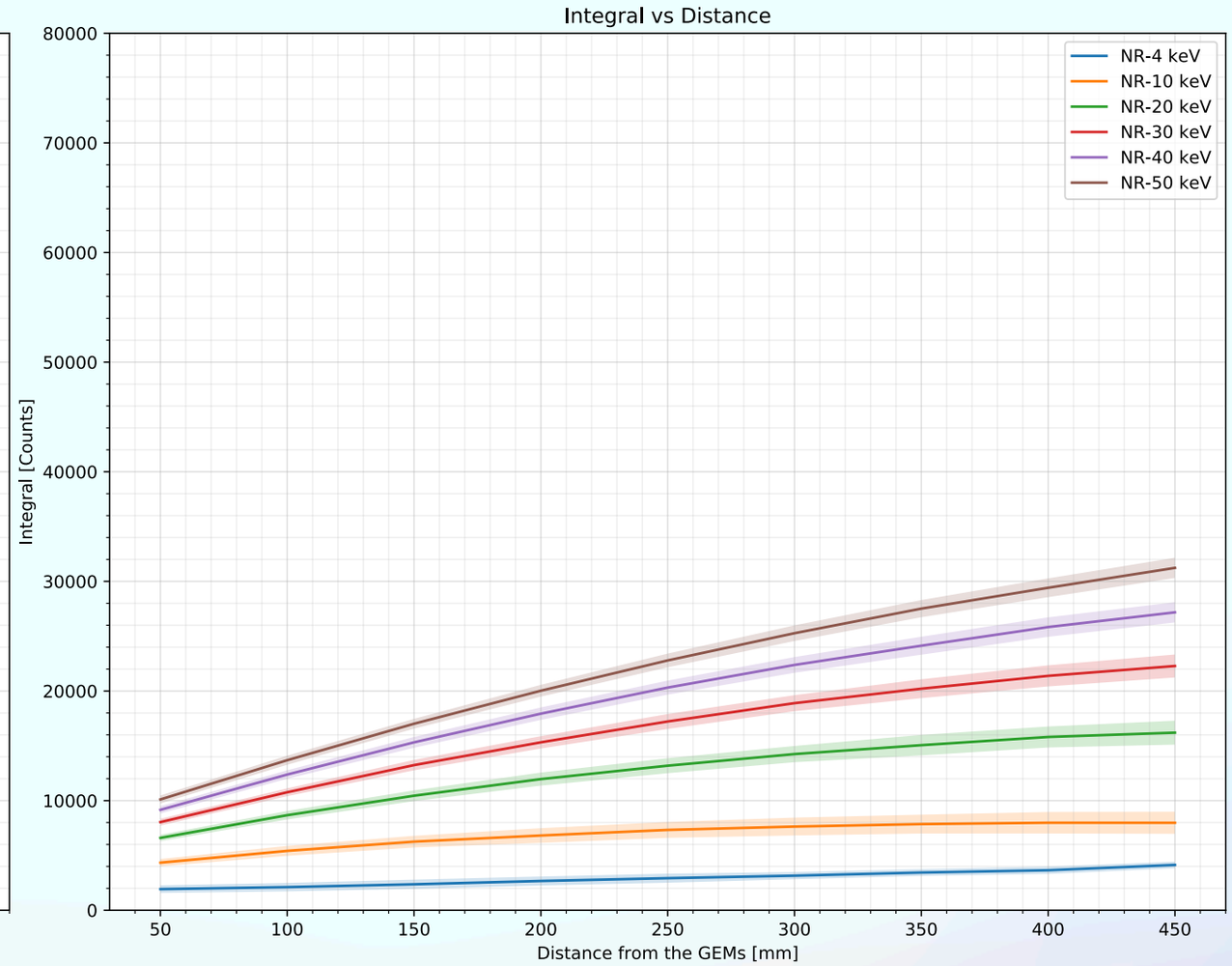




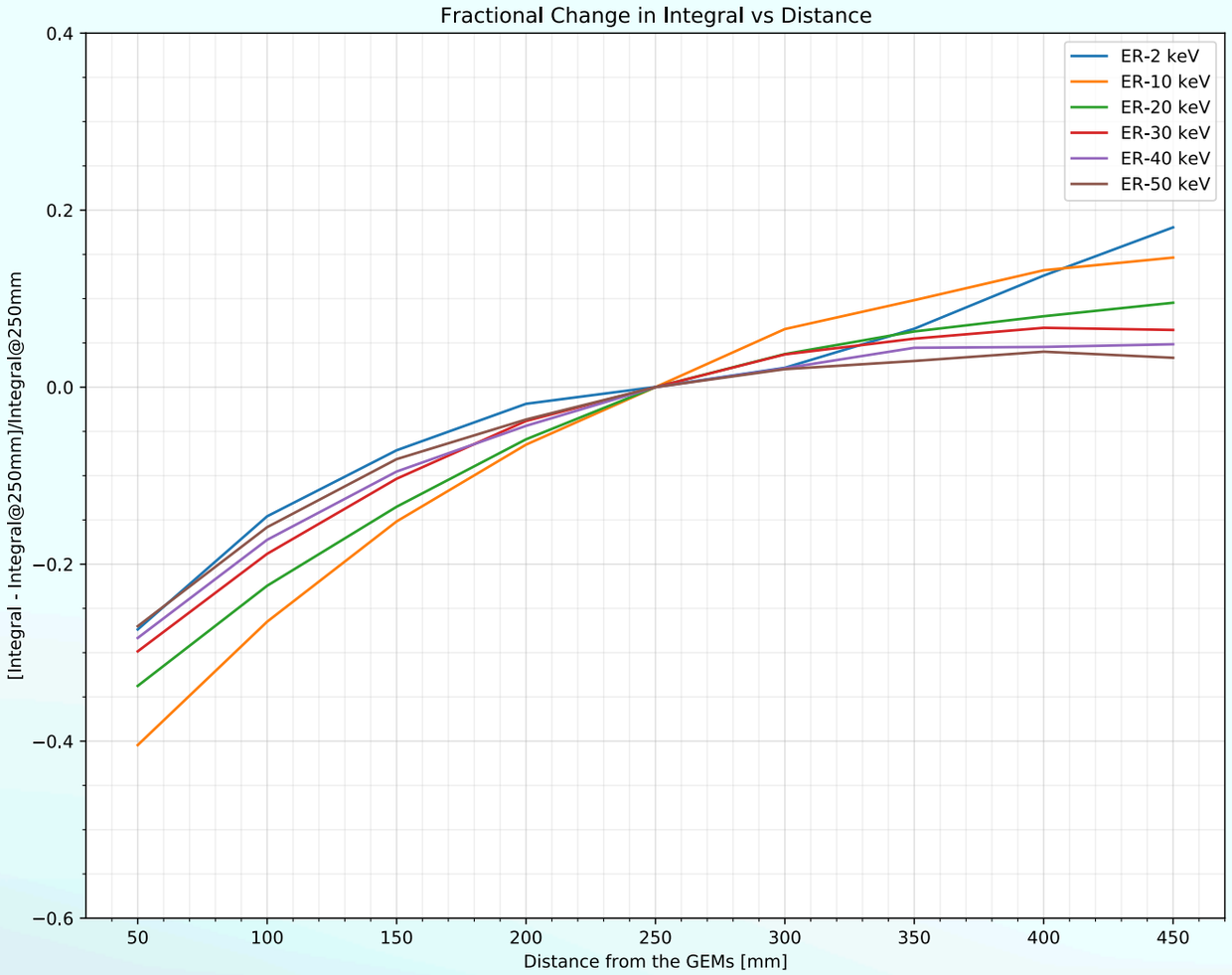
- Light integral ER data reaches a plateau after 25 cm while as light integral NR data increases as we go away from the GEMs at all the energies.

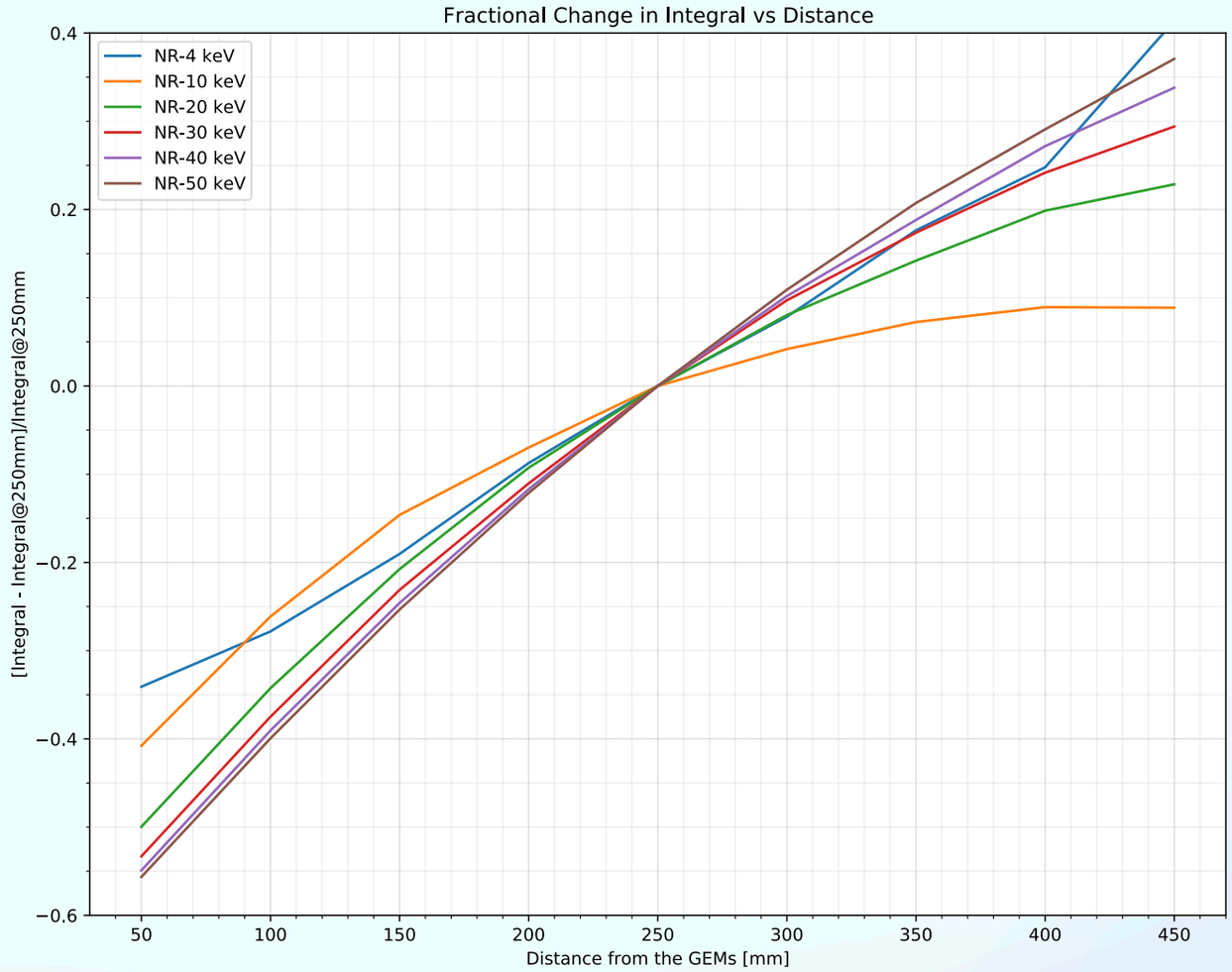
Integral vs Distance





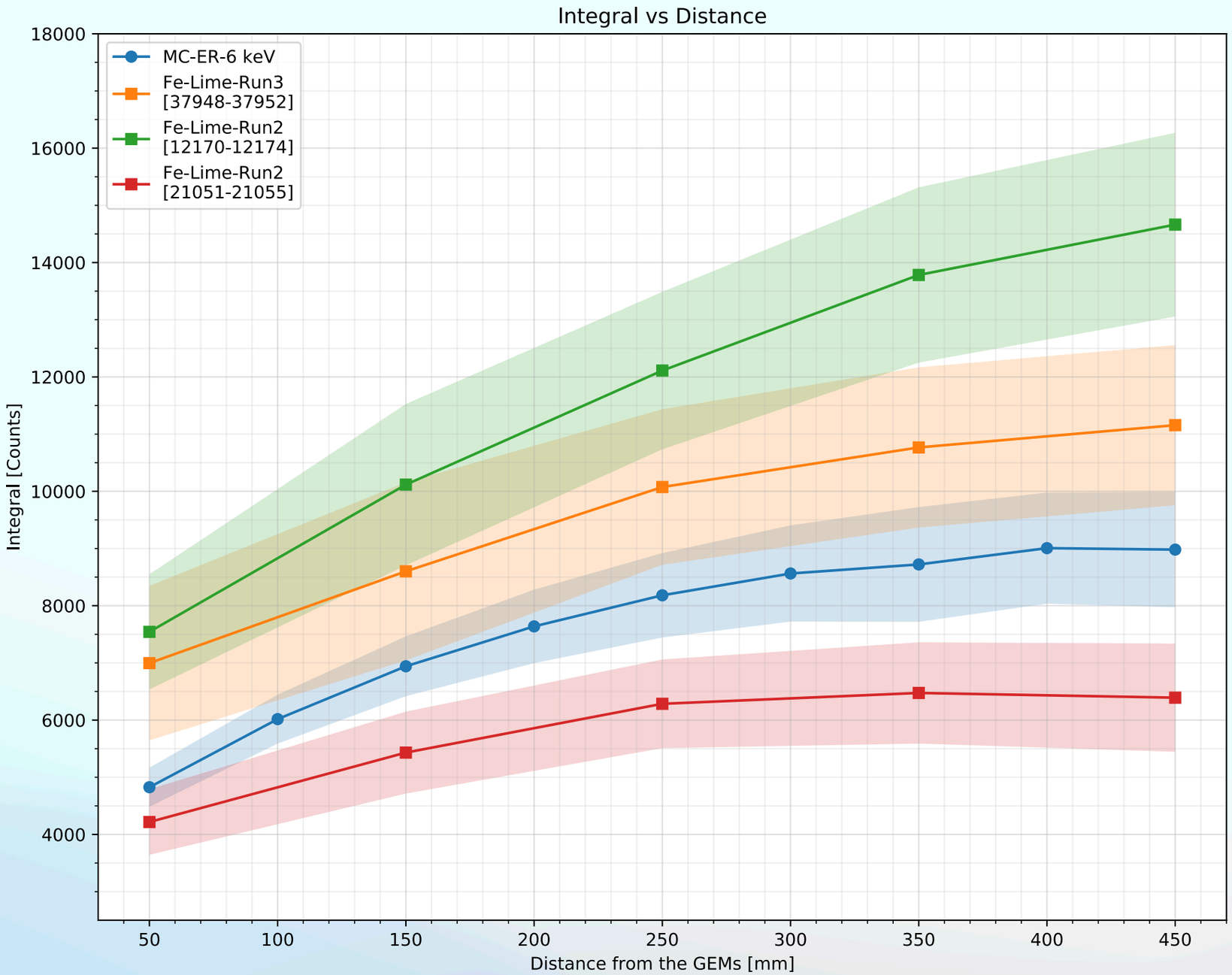
Fractional Change in Integral vs Distance





- Change in light integral for ER at all the energies with respect to the light integral at 25 cm away from the GEM is between ~[-30, 10]%. While for NR, the change is between ~[-55, 30]%.

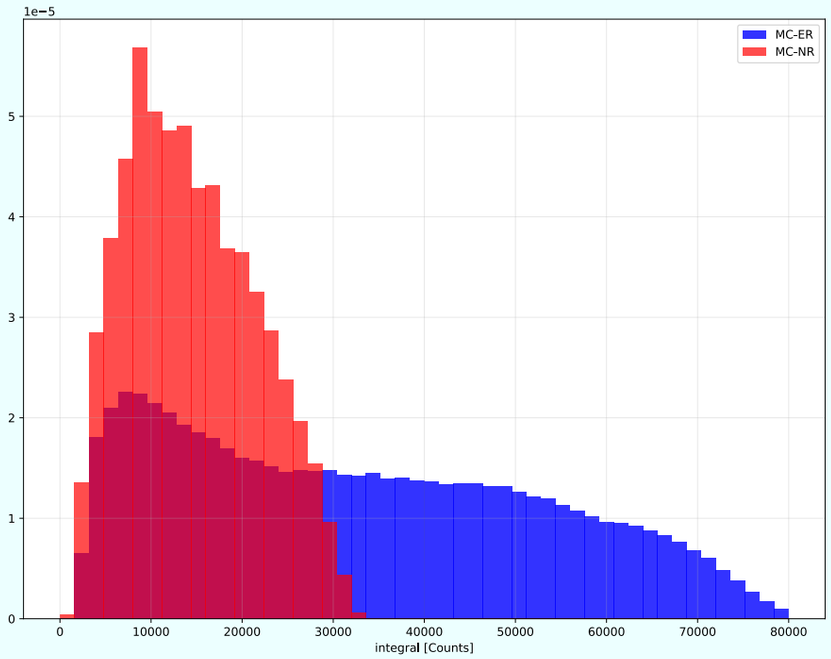
Integral vs Distance [MC/Data]

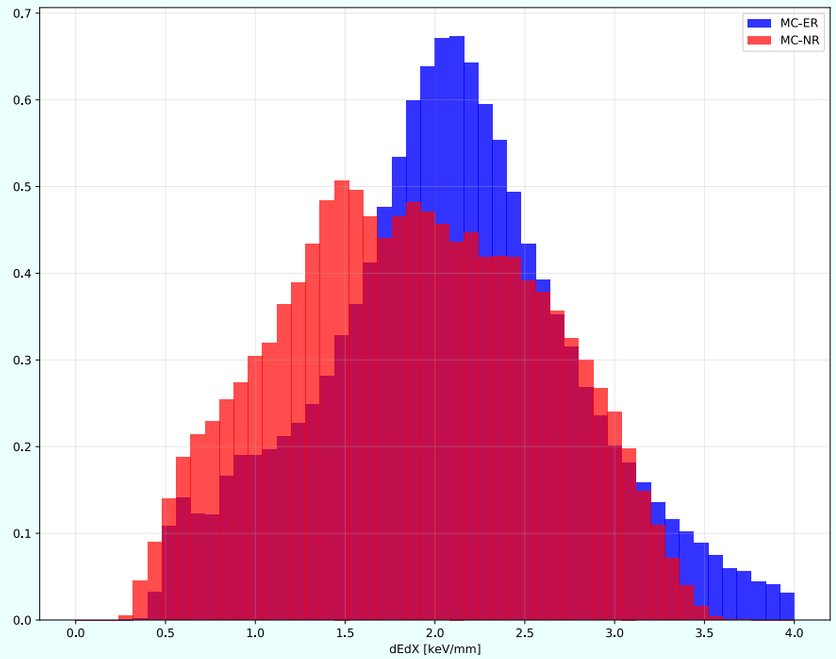


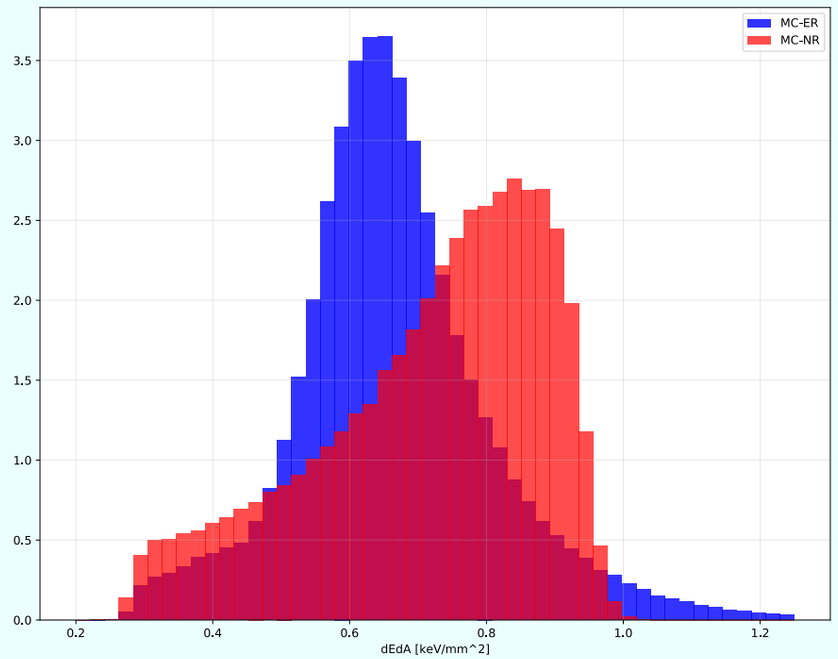
- Light Integral comparison of MC at 6 keV and Fe calibration data different runs. All the data sets have different light yield, so need to be corrected to match the MC.

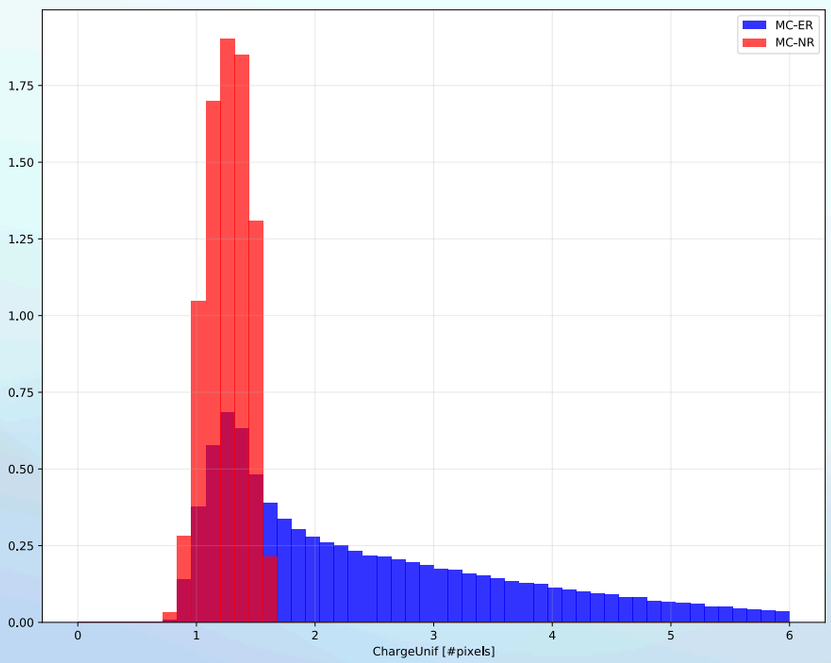
ER/NR Discrimination

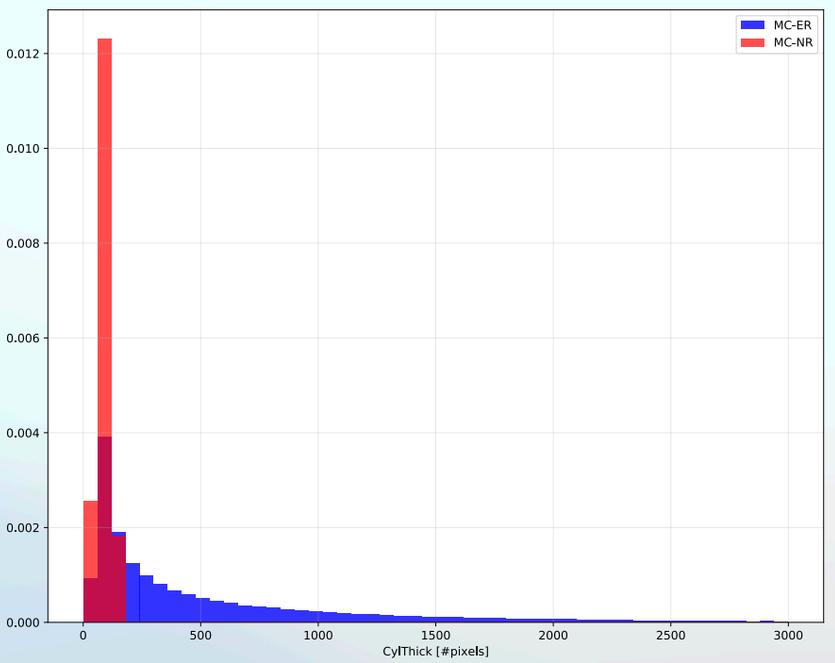
Histograms for Discriminating Variables

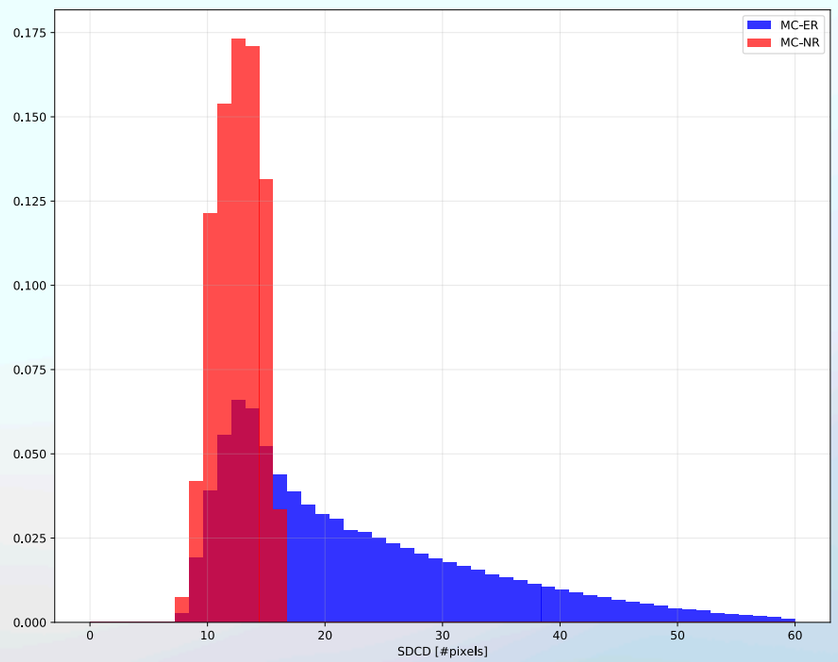




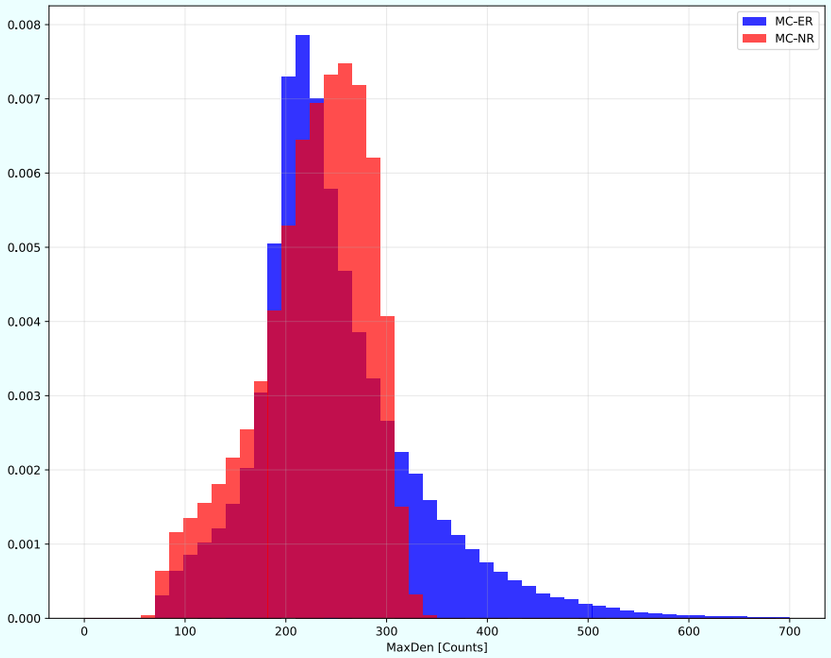


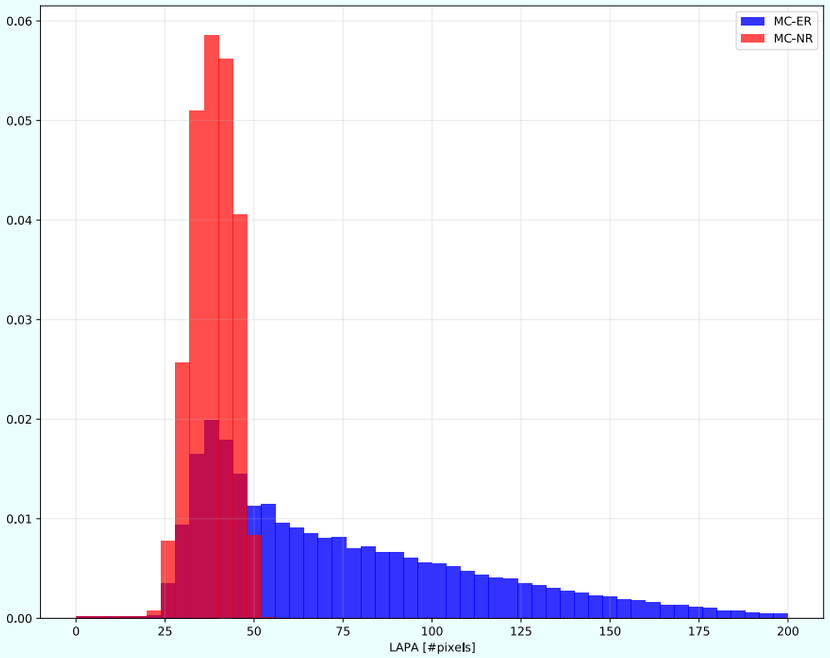


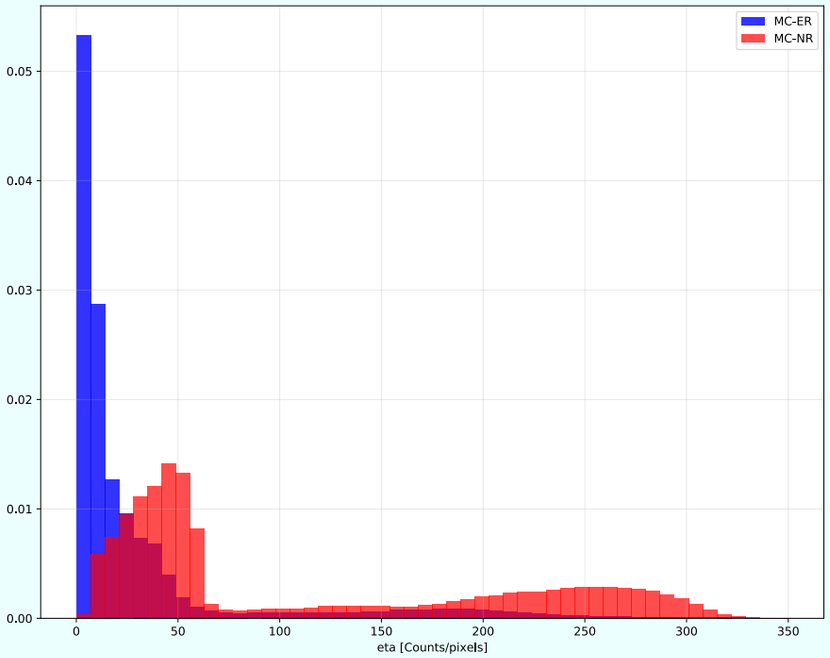


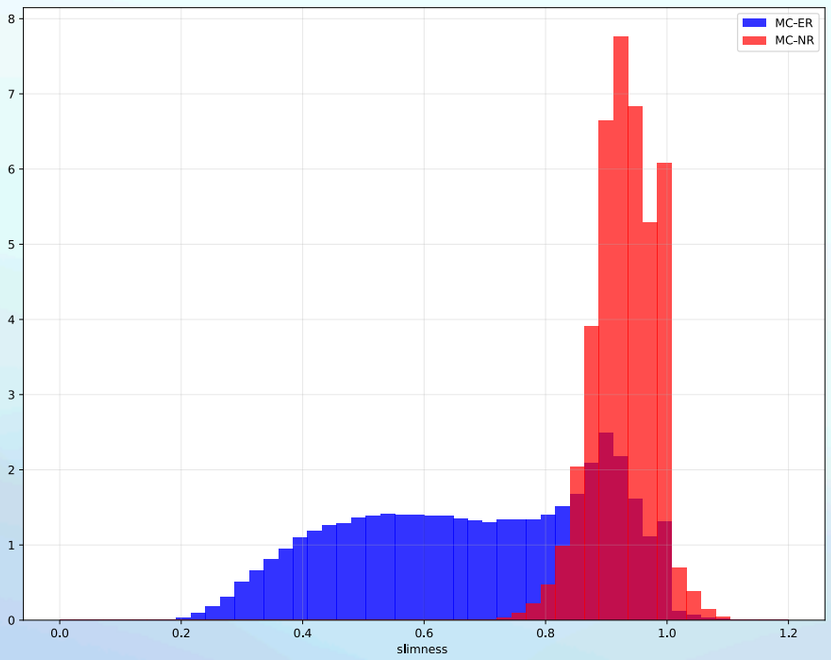


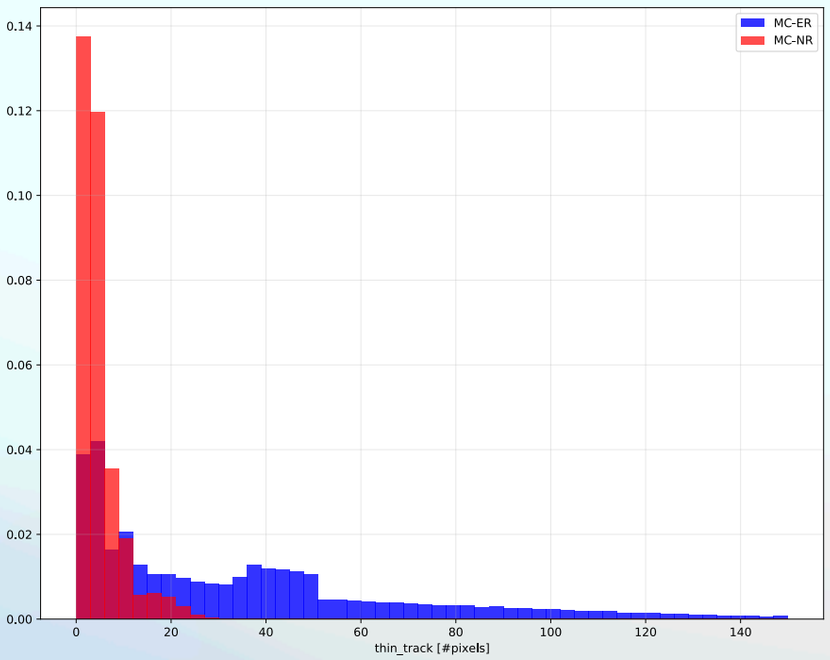
Histograms for Discriminating Variables

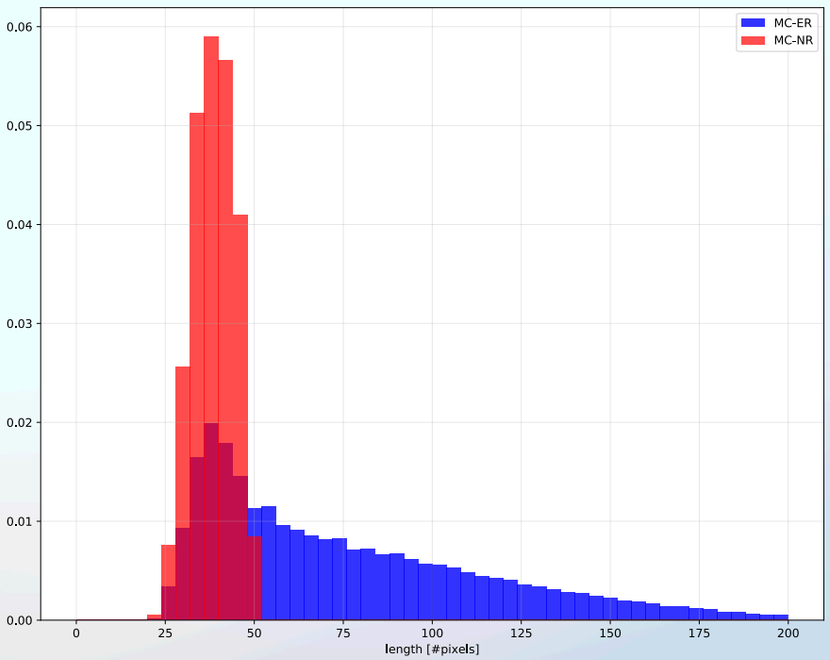




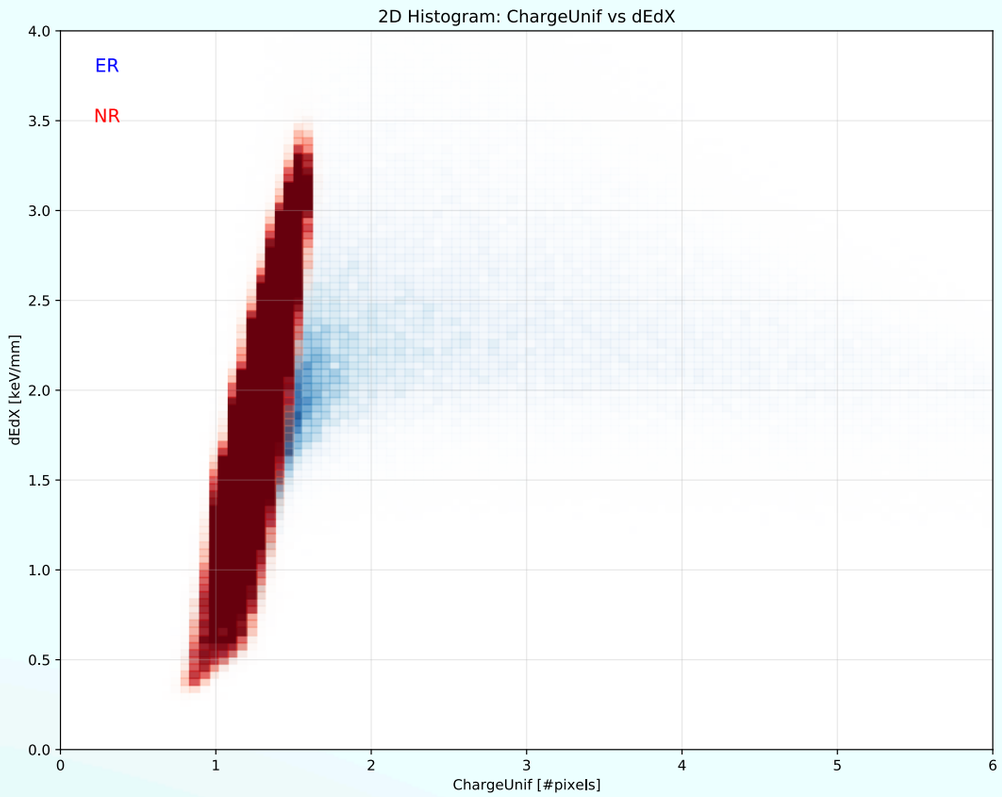


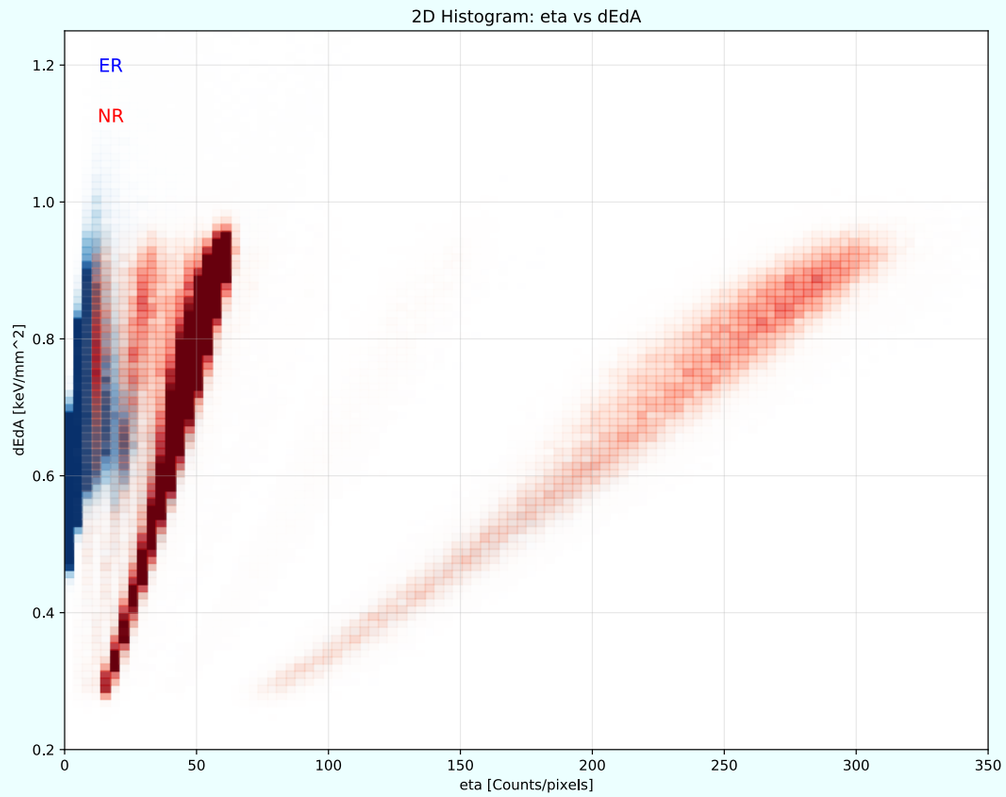


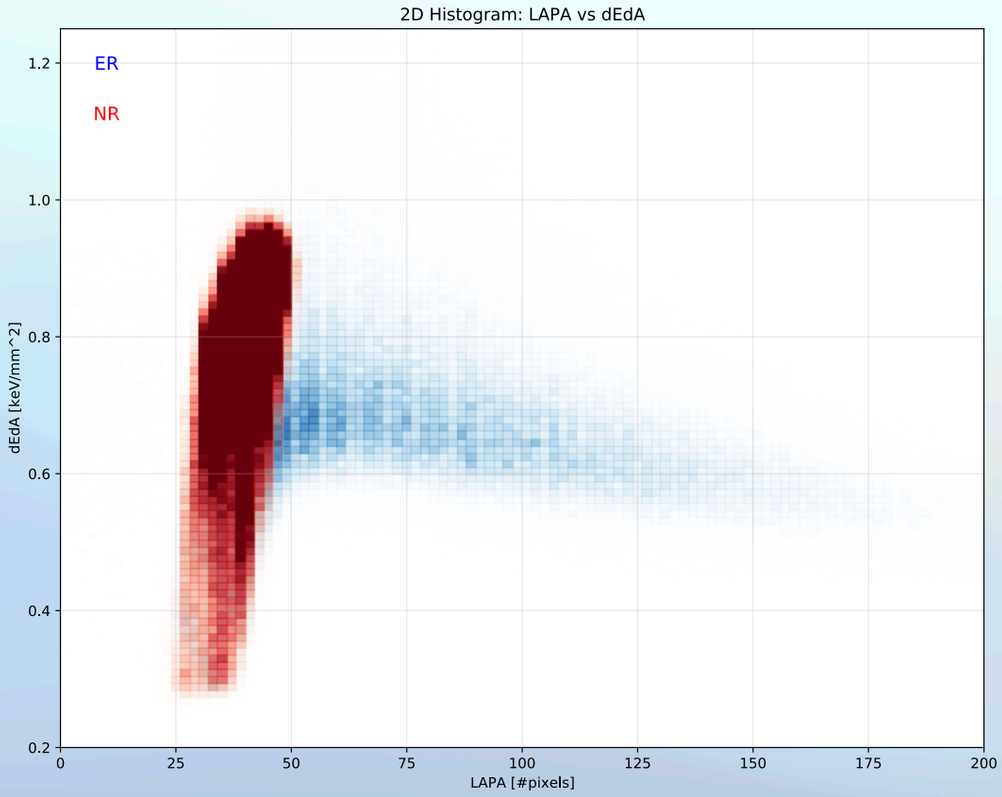


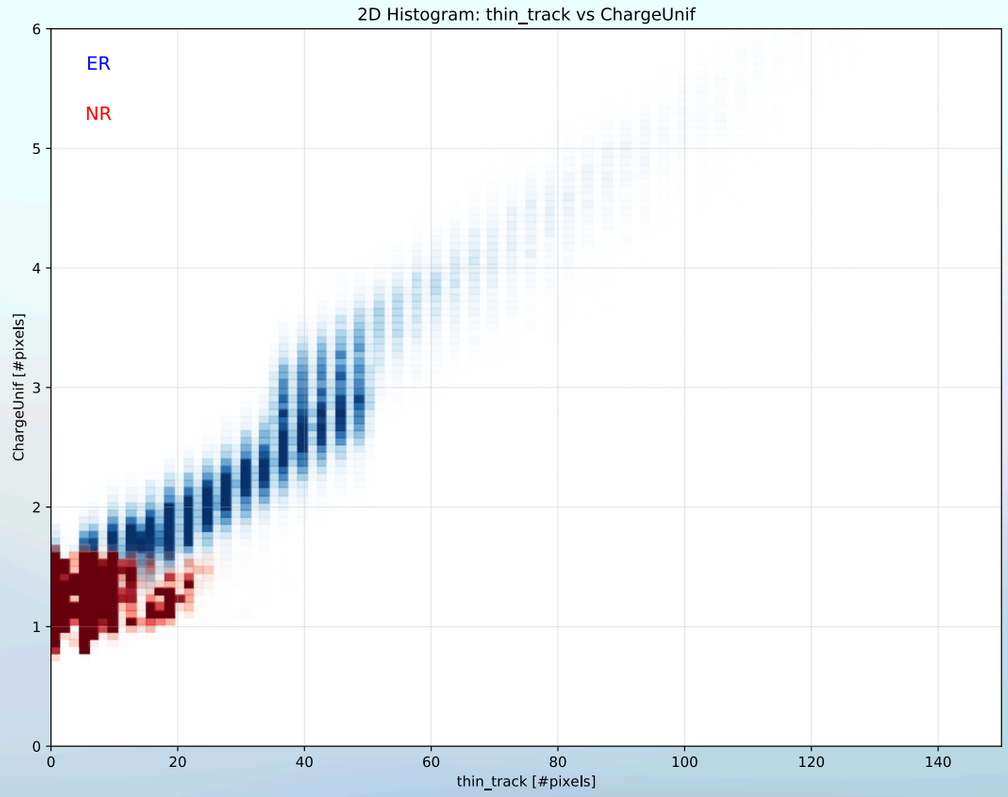


2D- Histograms

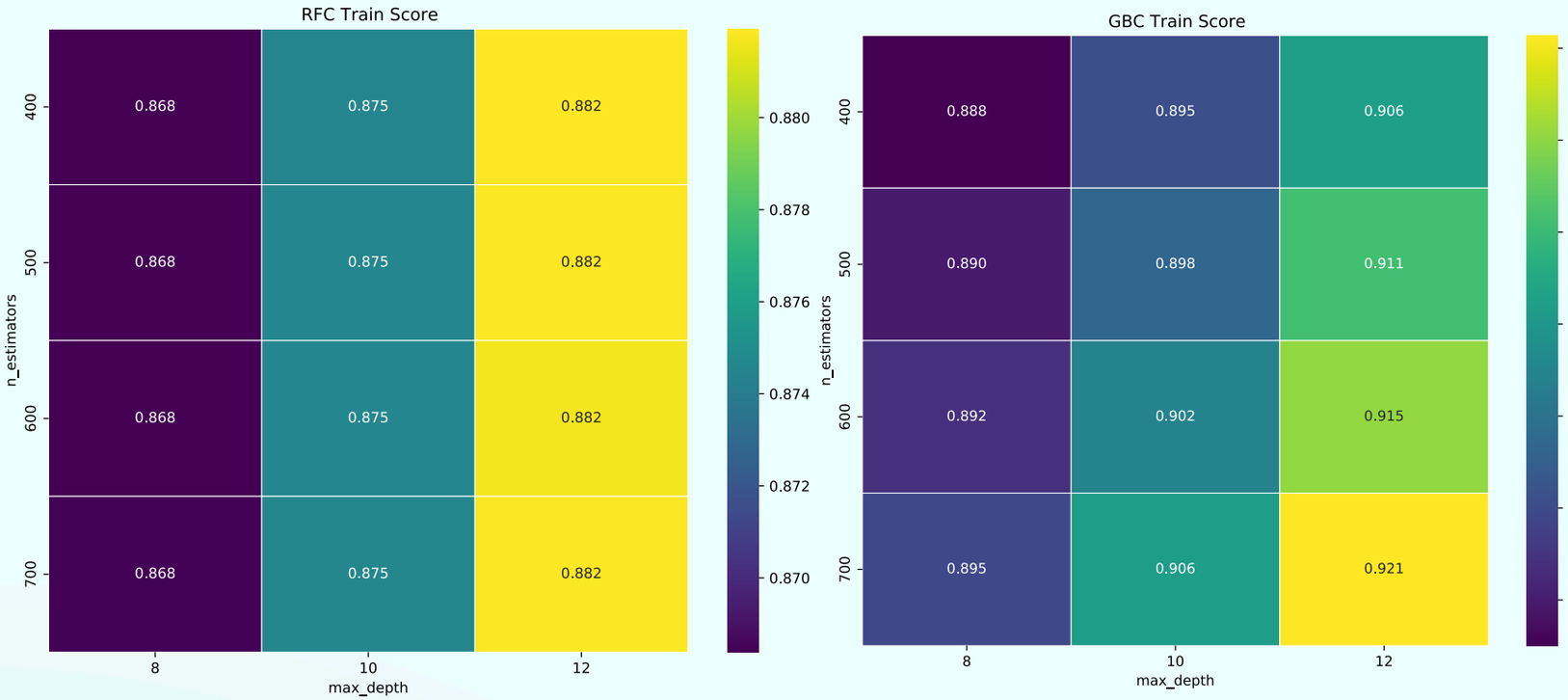


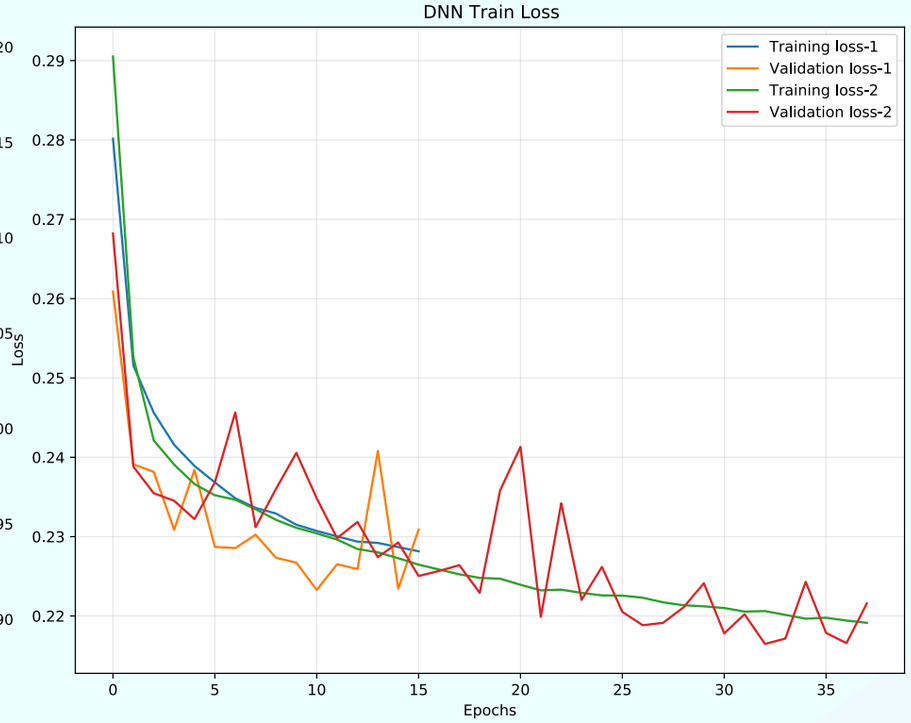






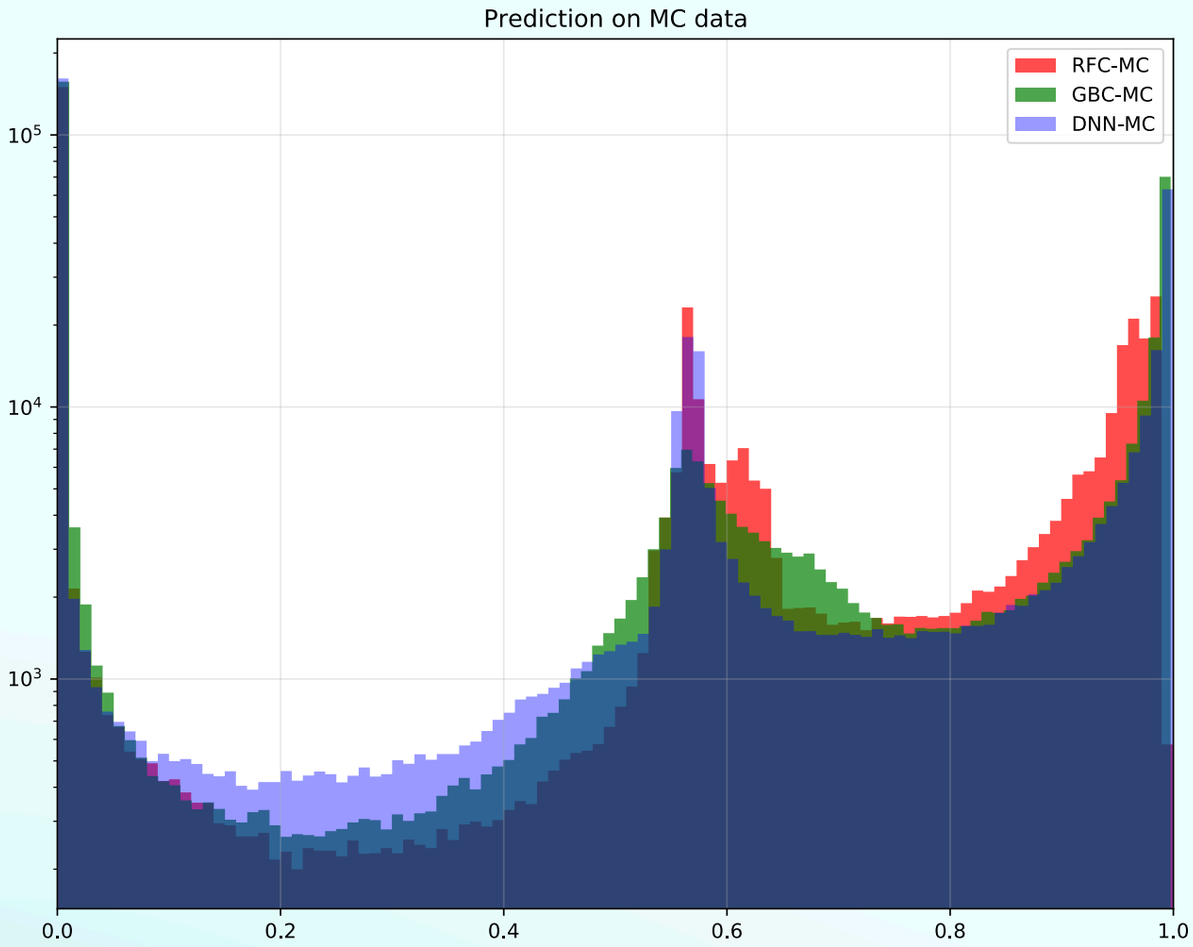
Training ML models

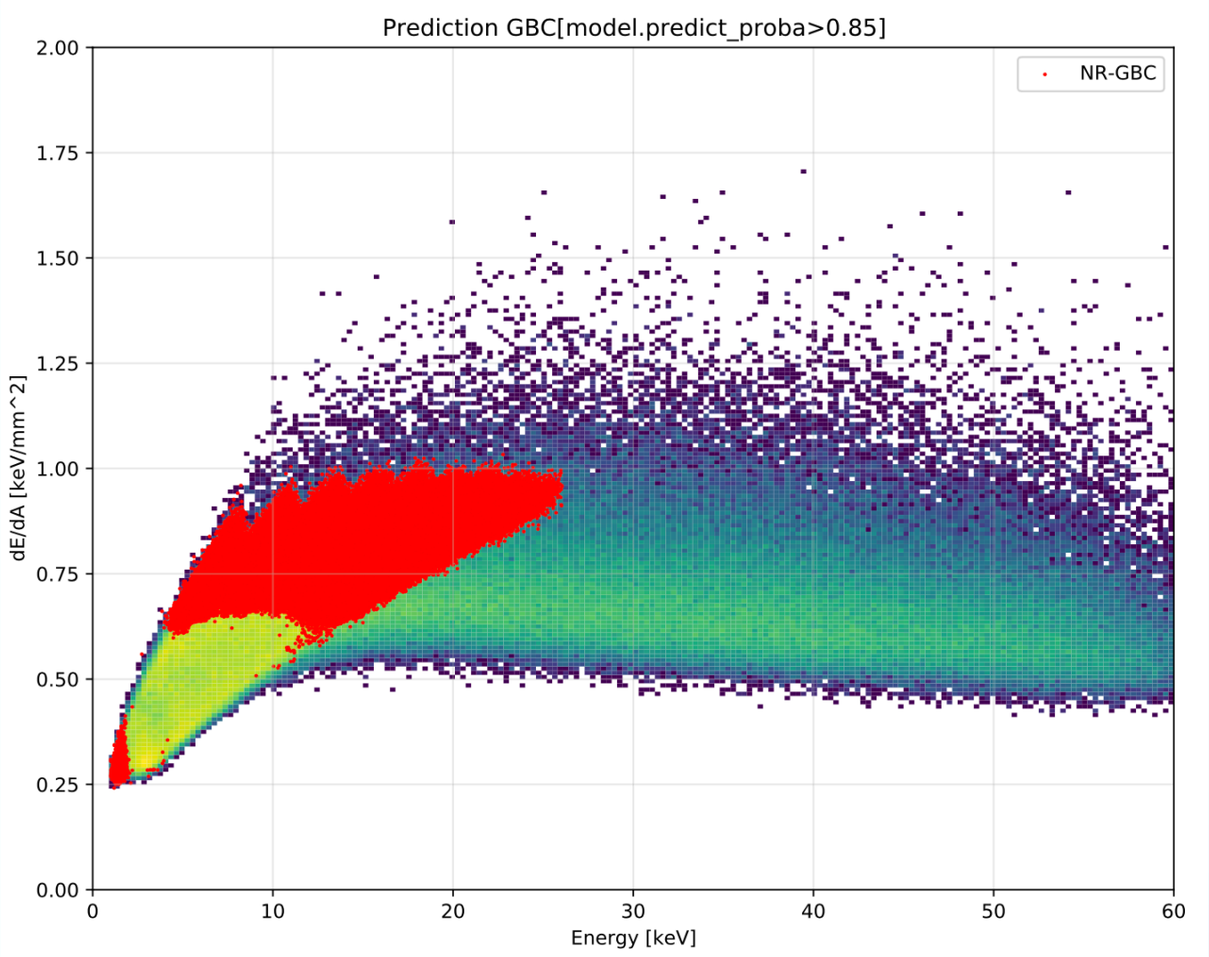




- Training Samples: 338319 [80% of the data]
- Validation: 84580 [20% of the data]
- Best parameters were chosen for the training by comparing the training and validation score.

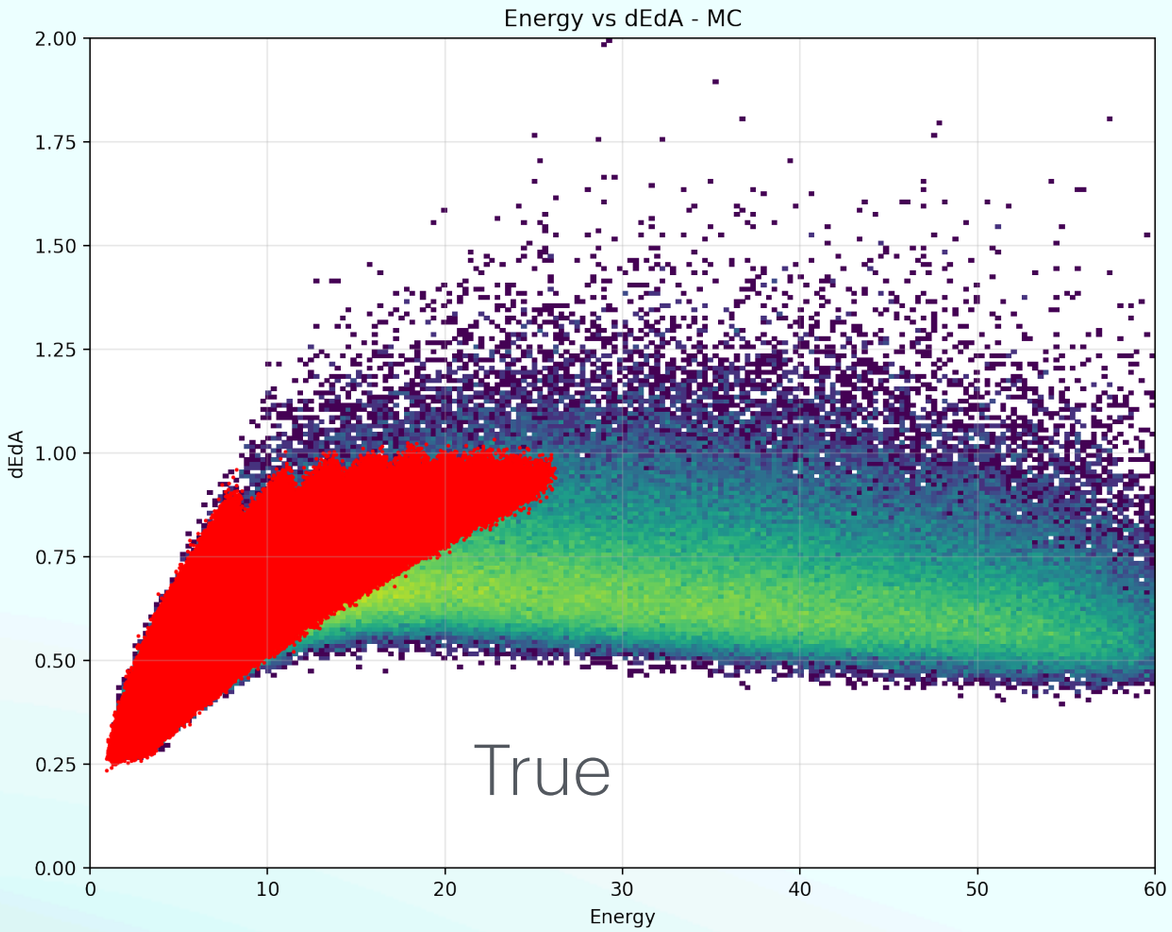
Prediction on MC

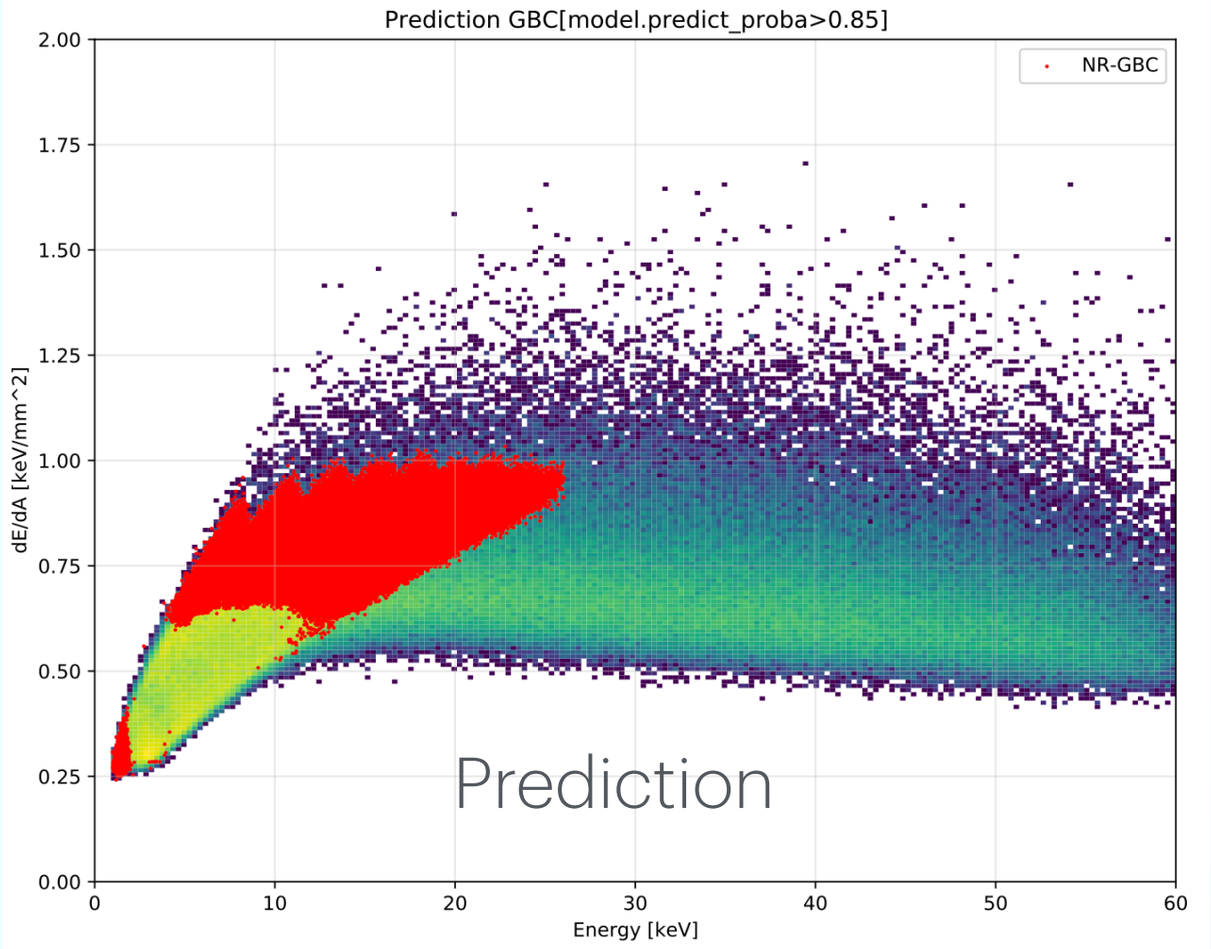




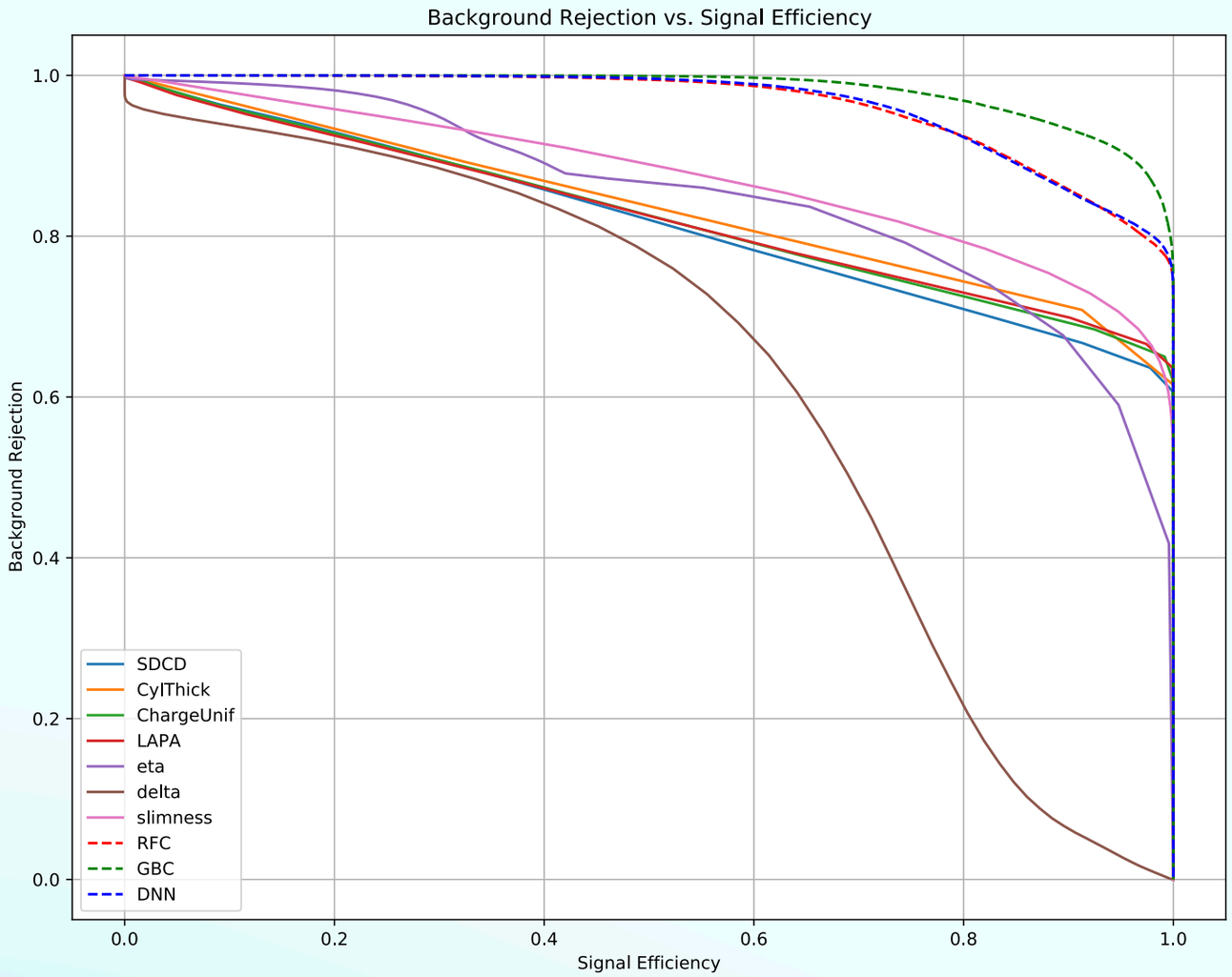
- Prediction is done with a threshold of 0.85.
- The bump in the middle of the histogram is because of the low energy region where ER and NR are overlapping and are very difficult to distinguish.

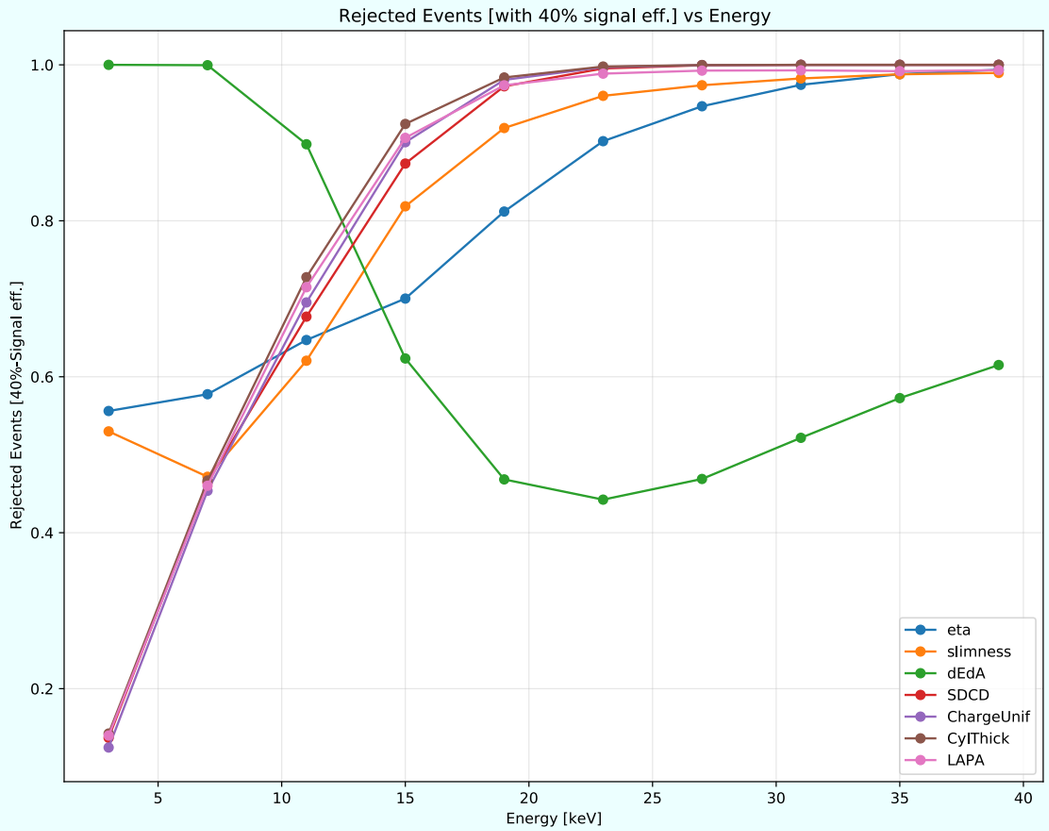
Prediction on MC

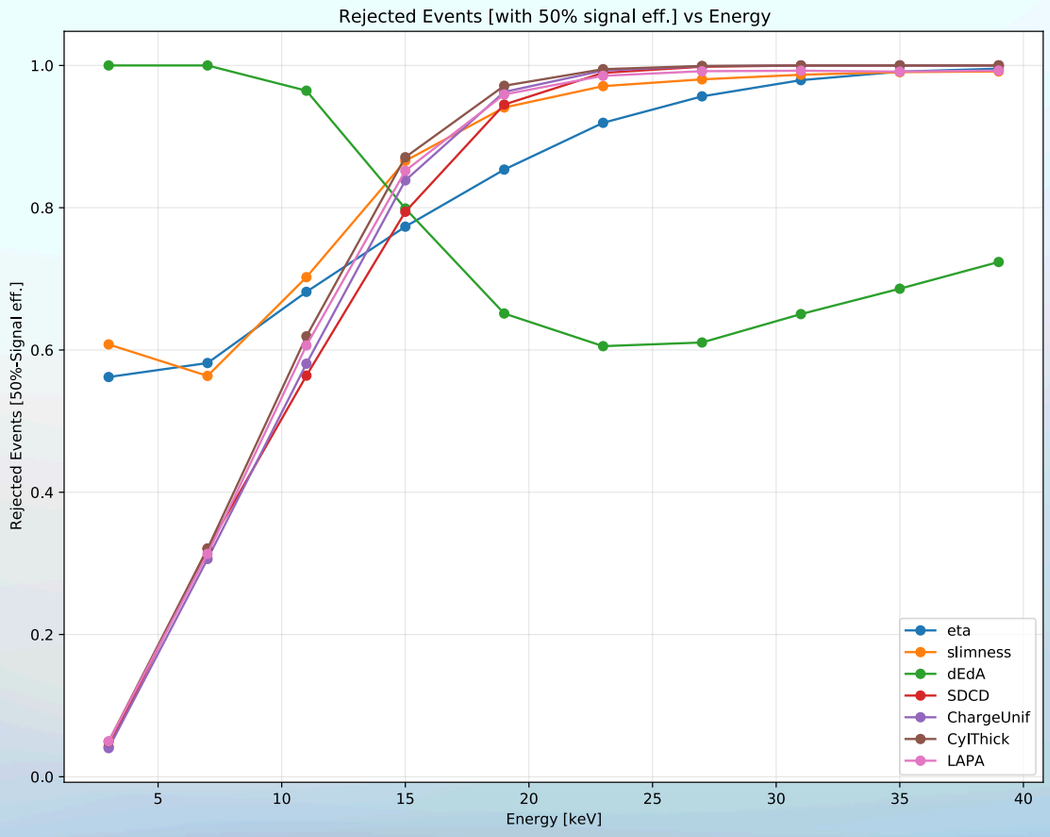




Bkg. Rejection

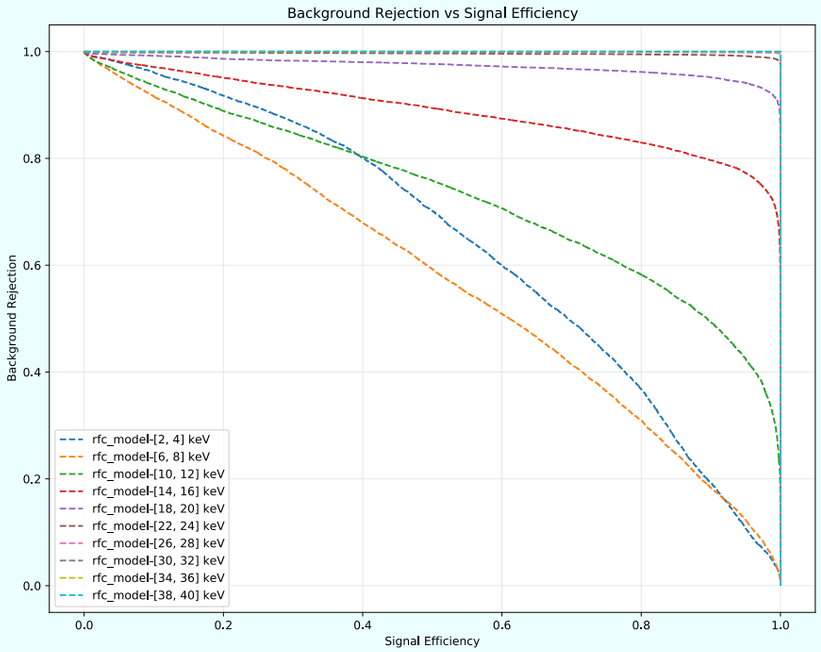


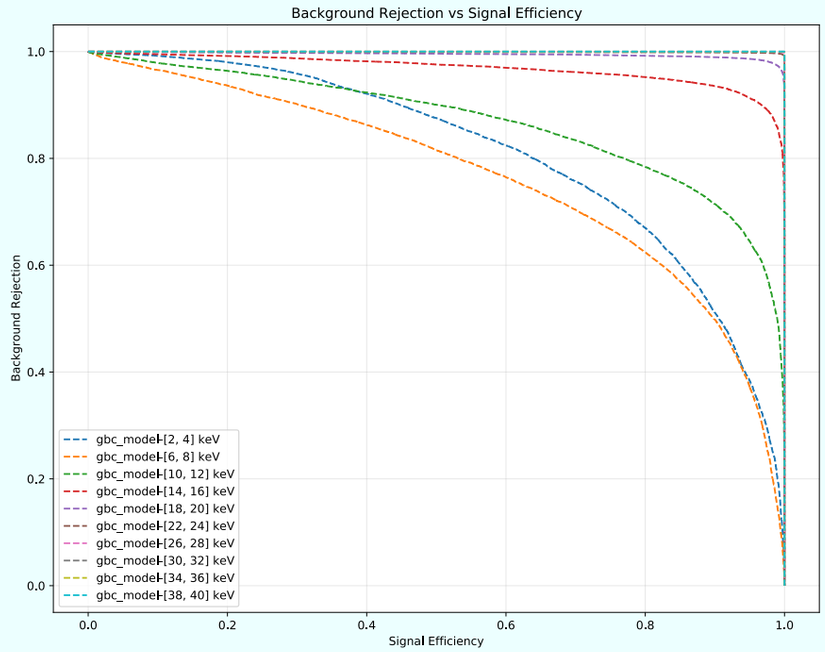


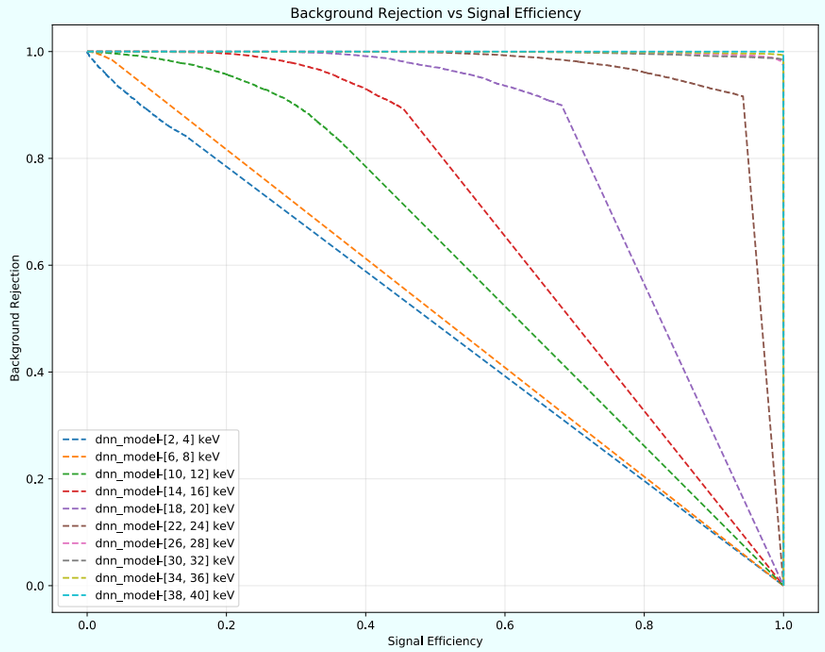


- First plot is Rejection vs signal efficiency on all the data.
- Second and third plots are rejection with the classical approach with 40 and 50% signal efficiency in each bin.

Bkg. Rejection in different energy bins ML

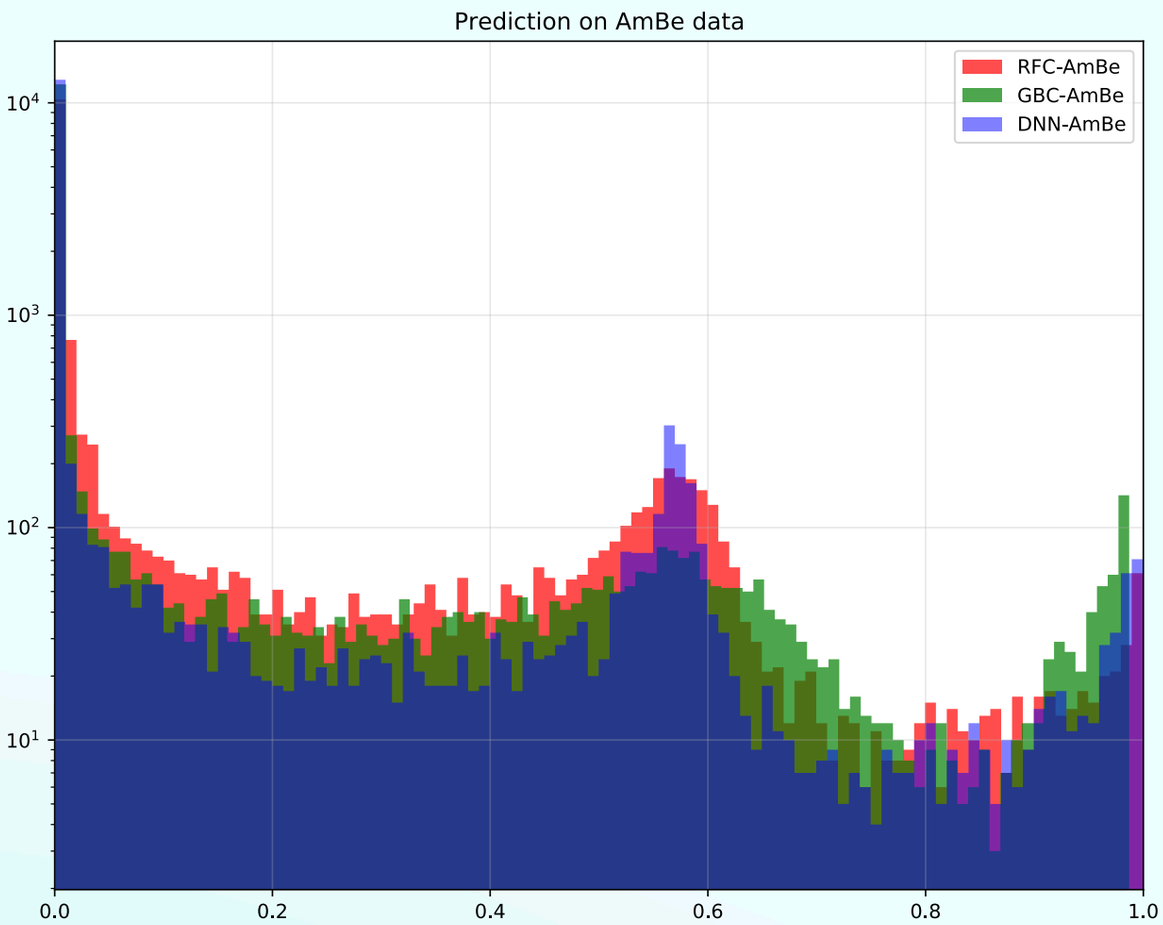


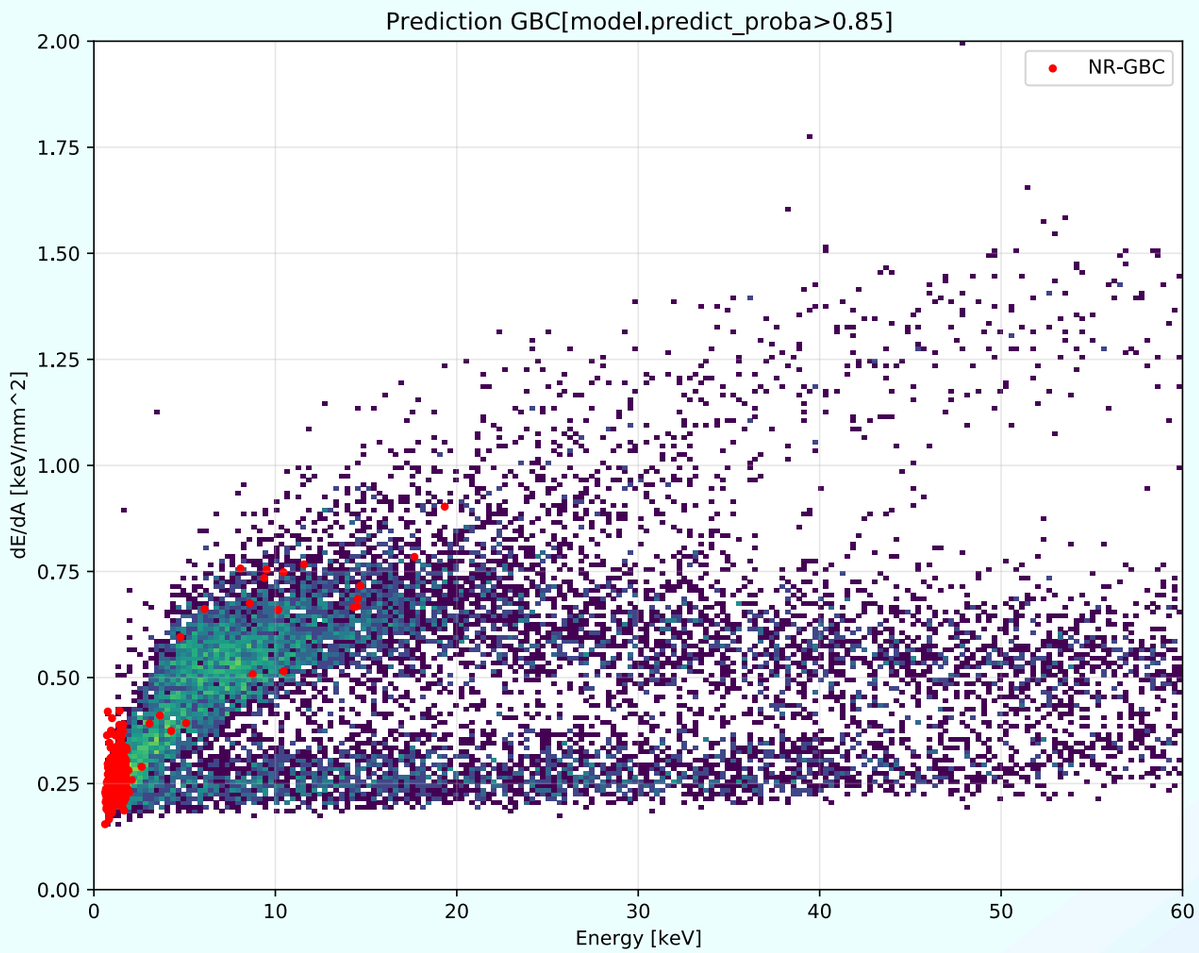




- Plots shows rejection using ML algorithms in the energy bins mentions on the plot.
- Here background rejection equals to 1 is ~2 x 10e4 events.

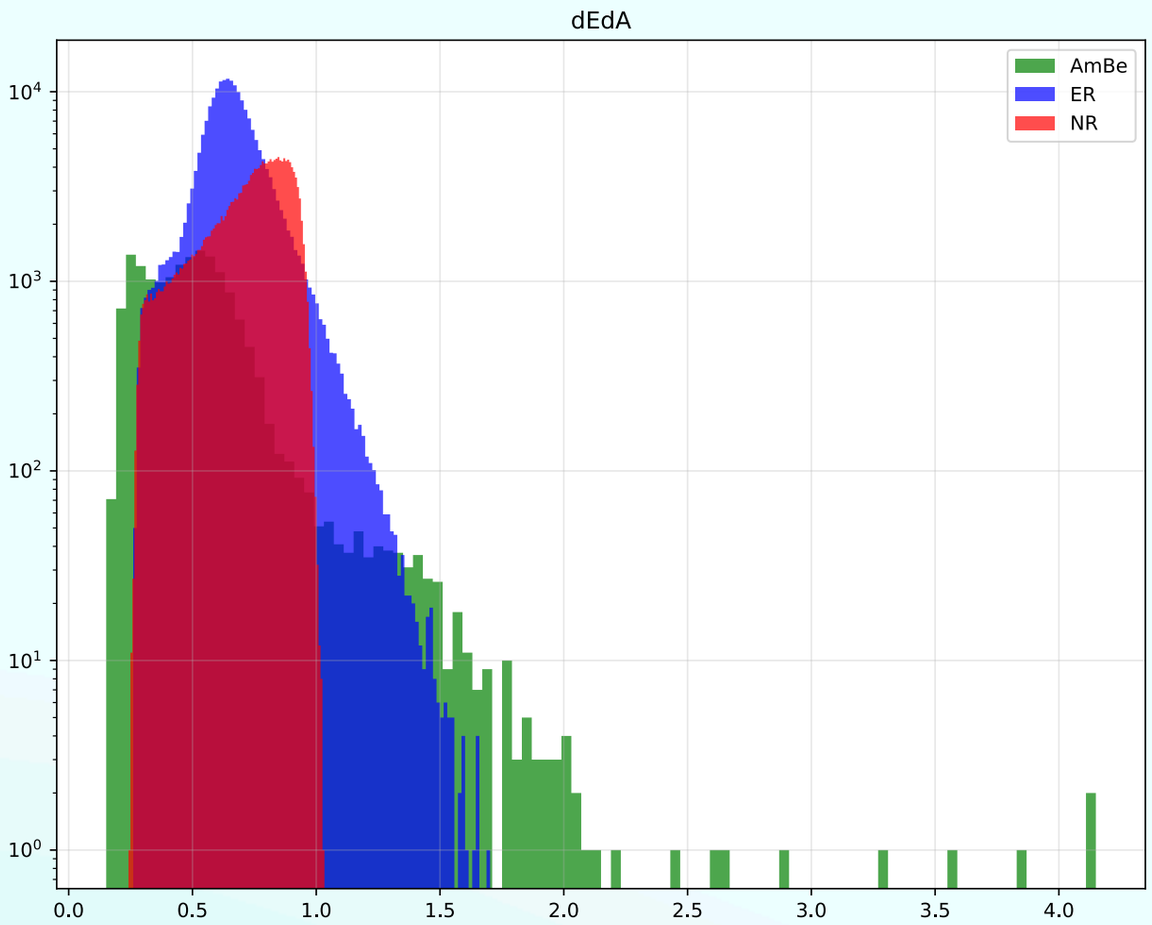
Testing ML models on AmBe data

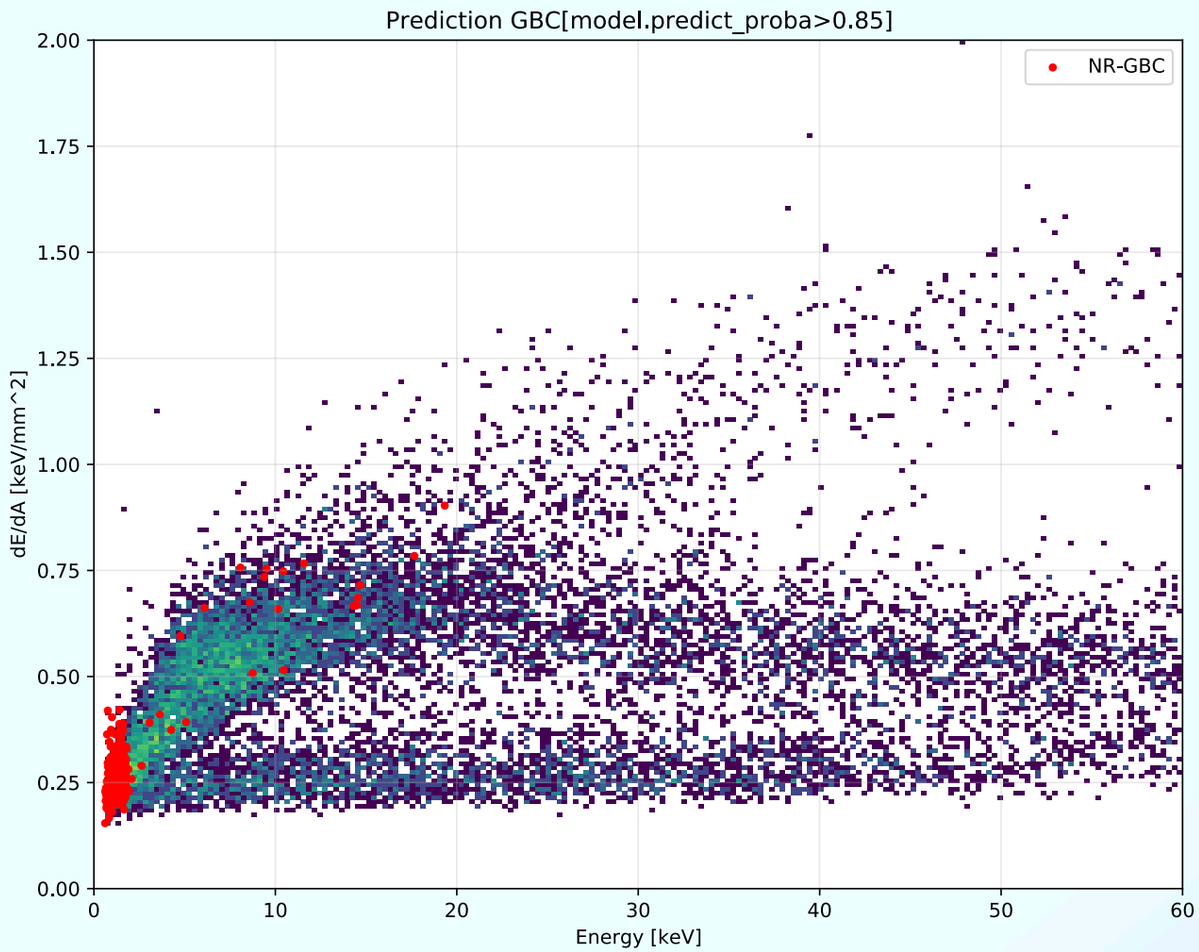




- Prediction on the AmBe LIME data is not good because the training data doesn't have the NR band that we see in the AmBe data

What's the issue?

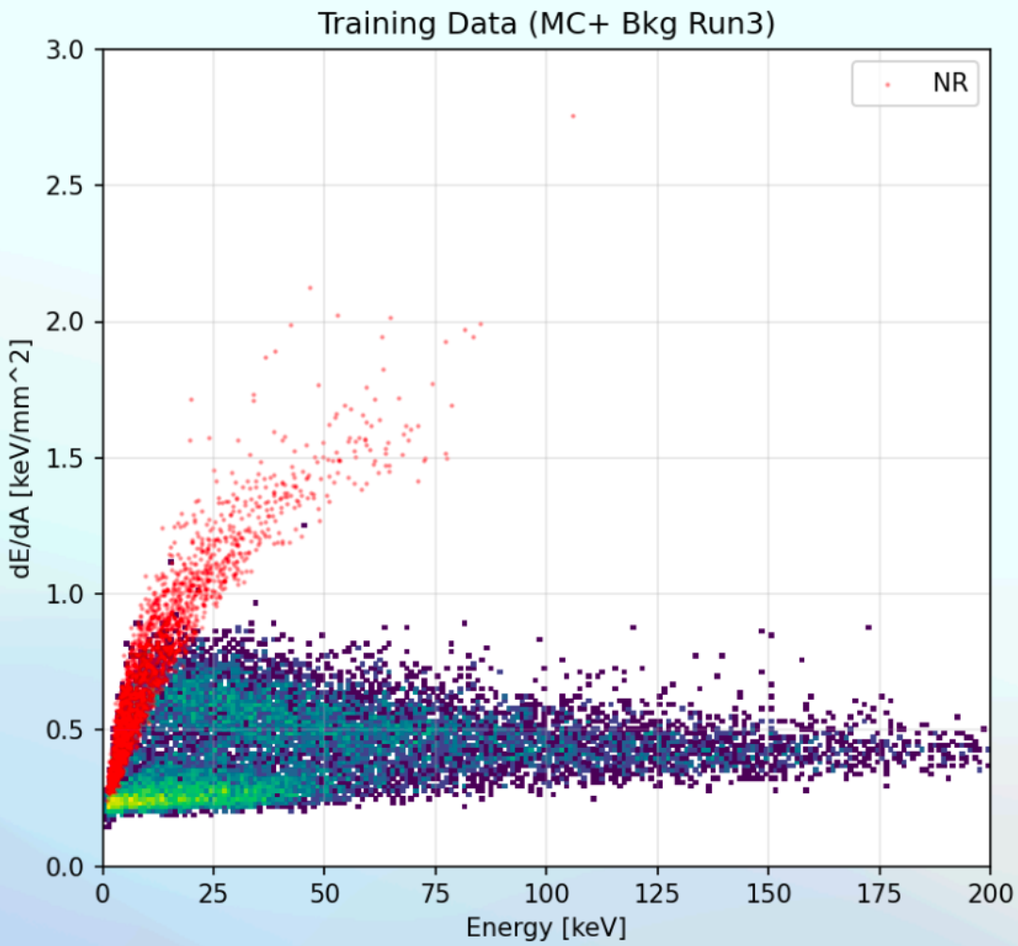


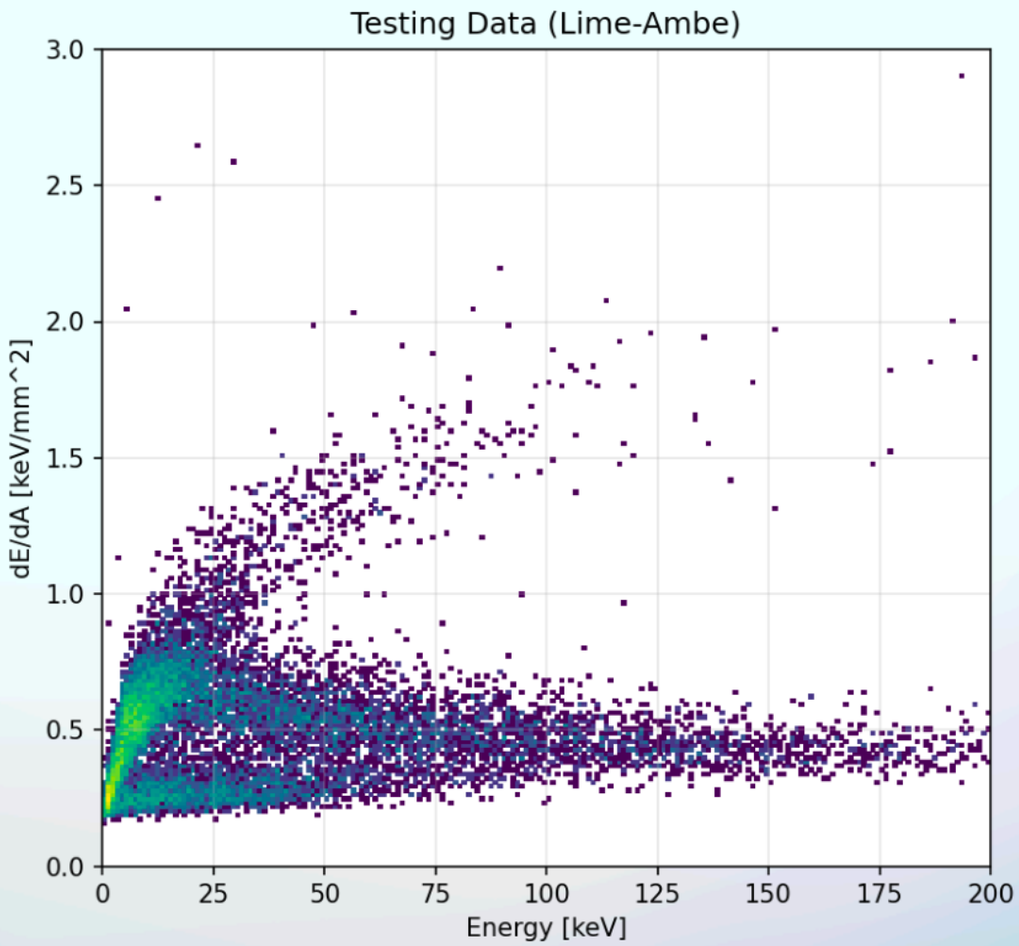


- ML algorithm are predicting the NR events in the range of data in which they are trained.
- Probably use of a semi-supervised learning can help or we modify our MC to match the data.

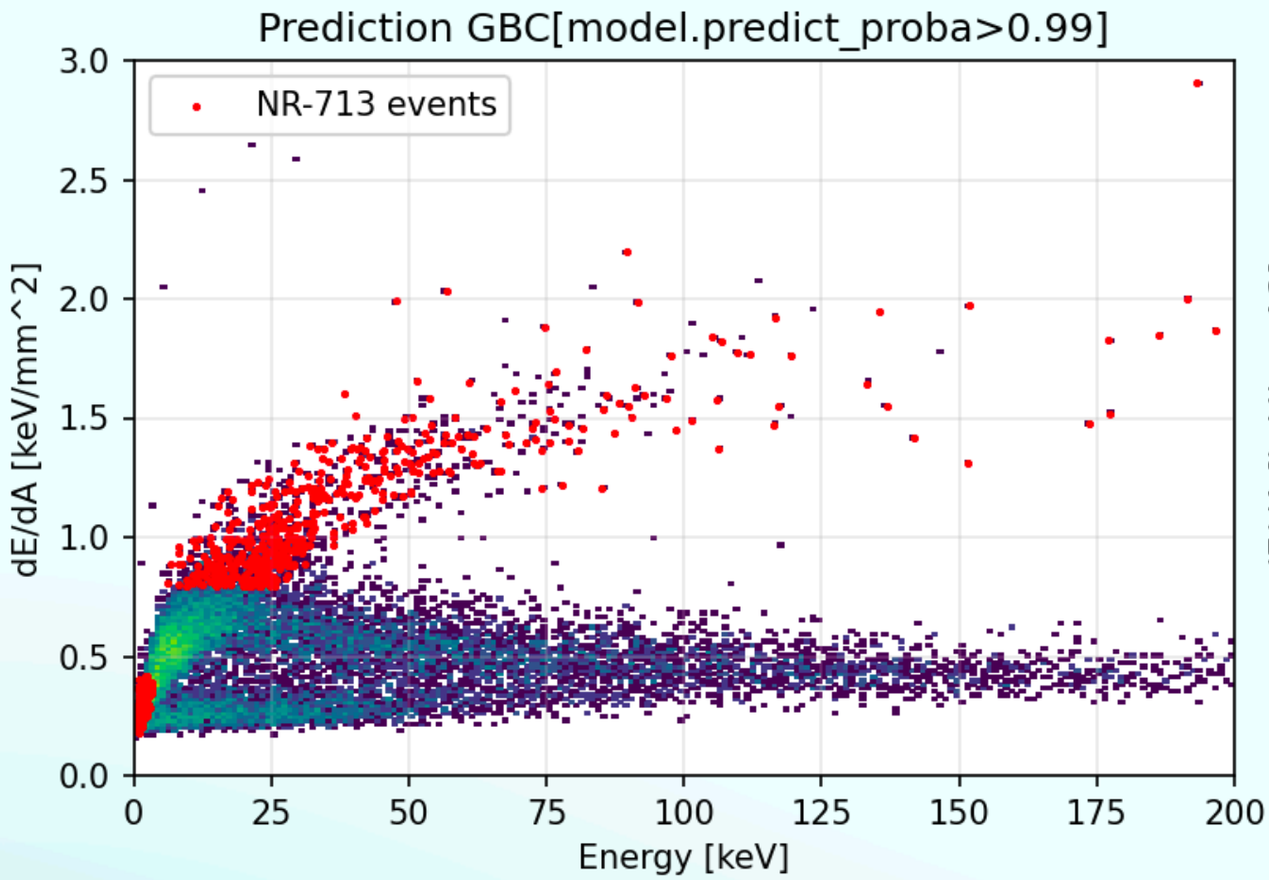
Old Analysis of AmBe data with ML

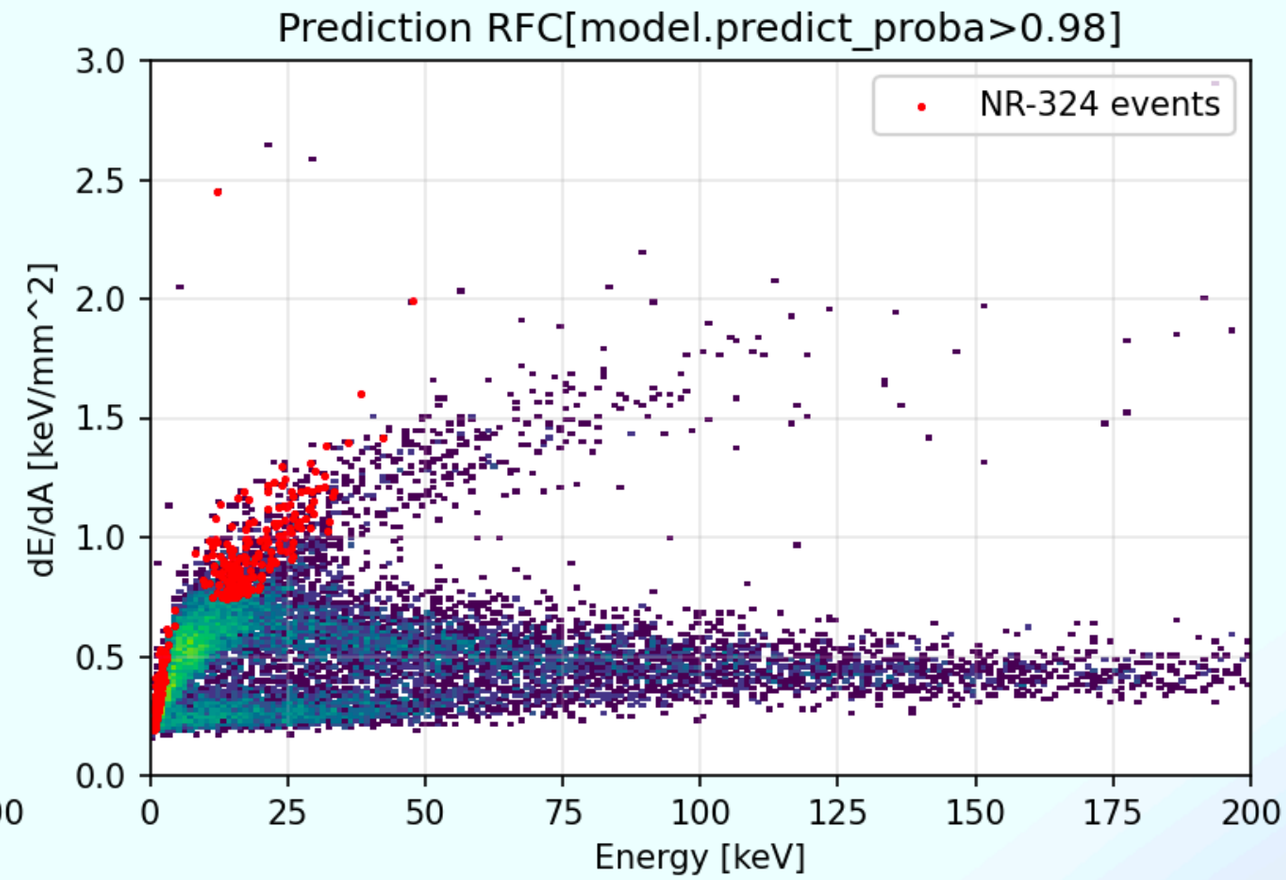
- Used AmBe simulations + BKG data [run3] for training.
  - Reason: expected NR events in BKG is ~1000/year
  - Bkg data selection: dEdA <0.7 and slimness < 0.8
- Variables: Energy, dEdA, dEdX, SDCD, CylThick, ChargeUnif, LAPA, MaxDen, eta, slimness, size, tgausssigma





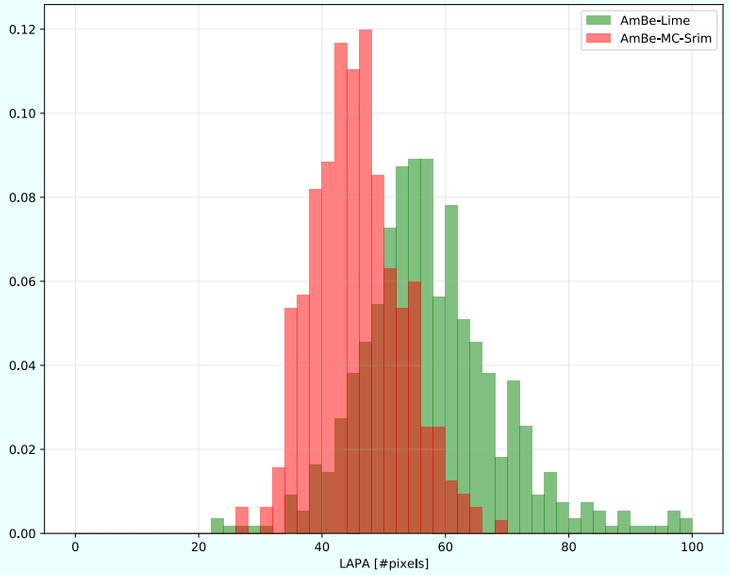
Prediction on the AmBe data

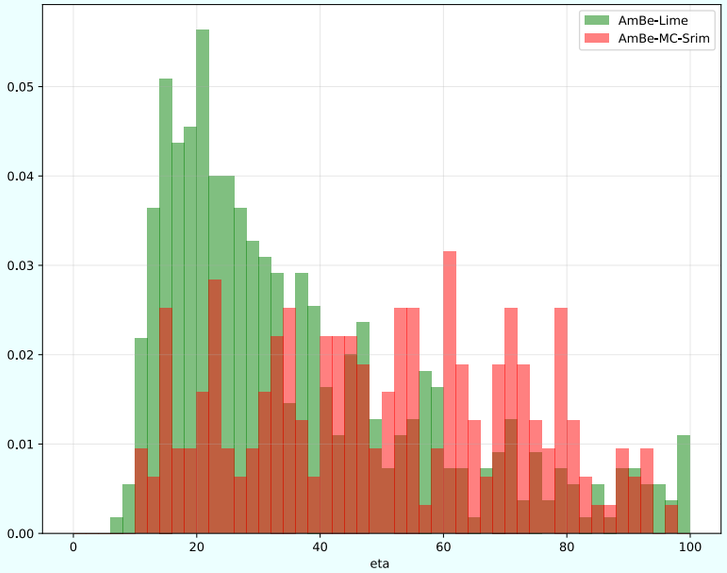


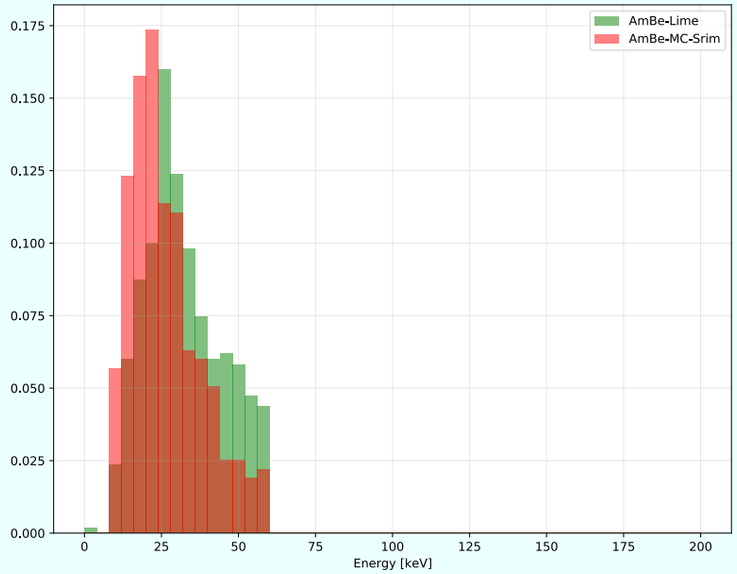


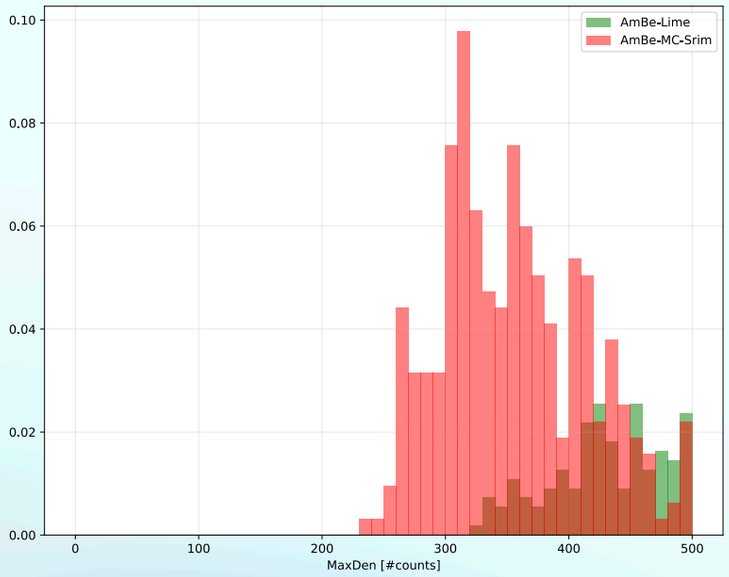
- Prediction on the AmBe data was correctly selecting the NR band, because NR band was present also in the training data.

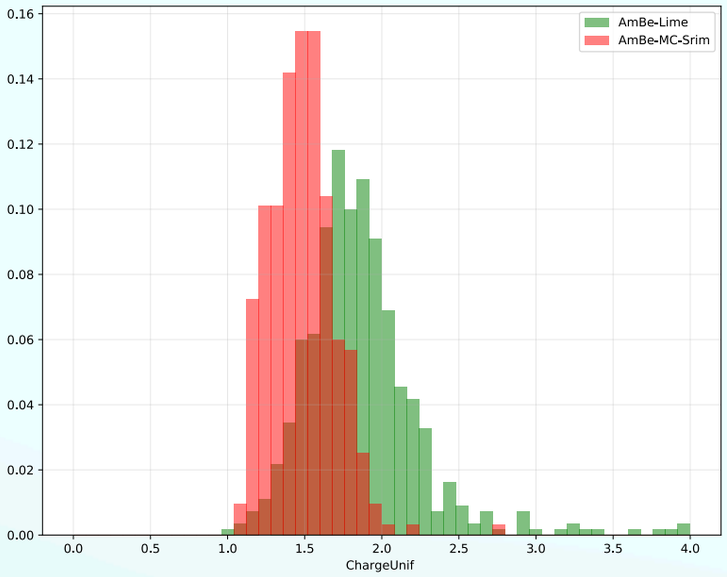
Shape variables comparison-NR

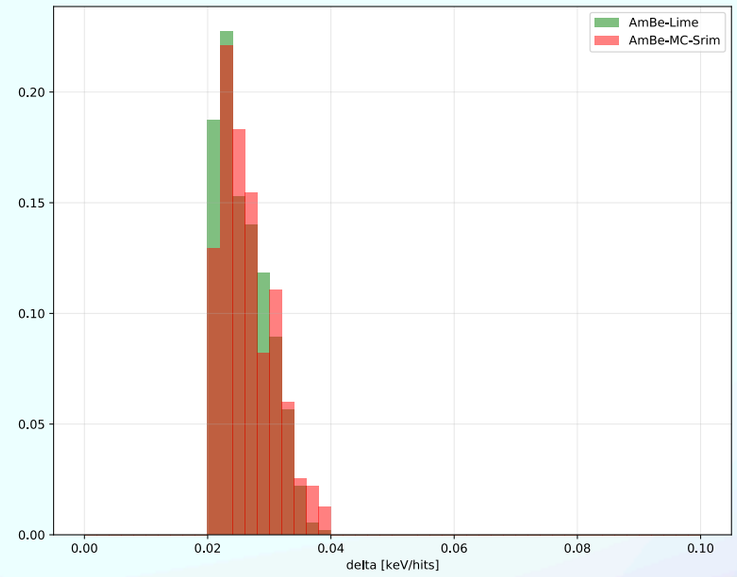


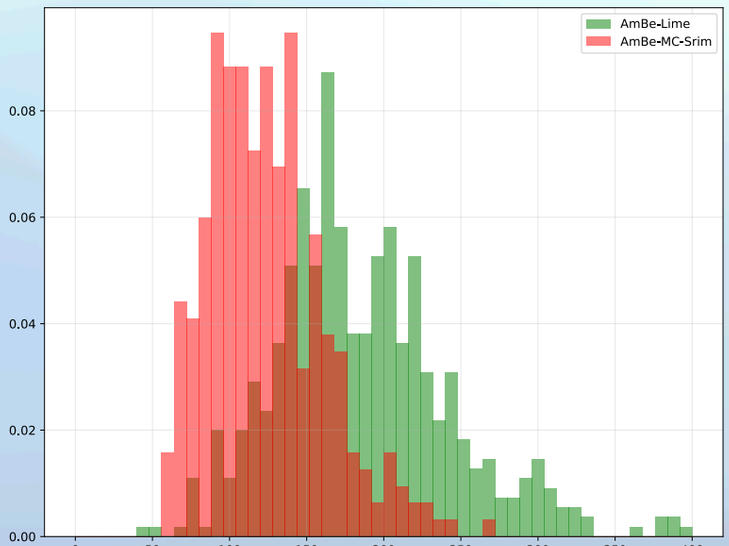


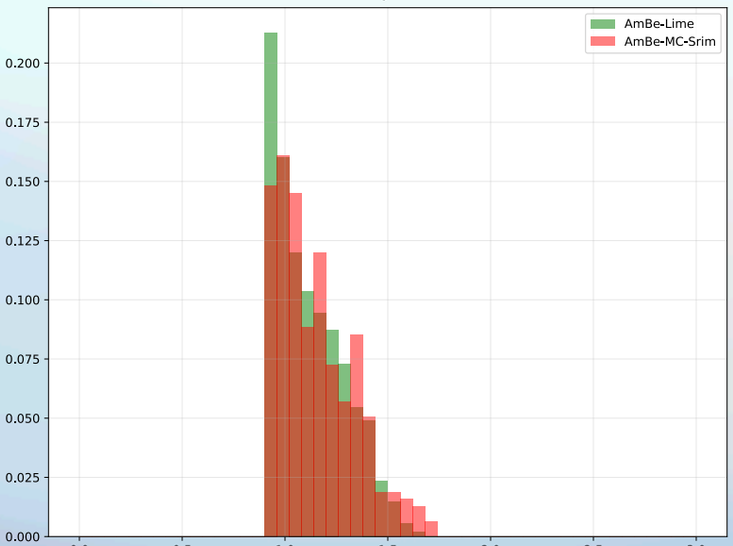






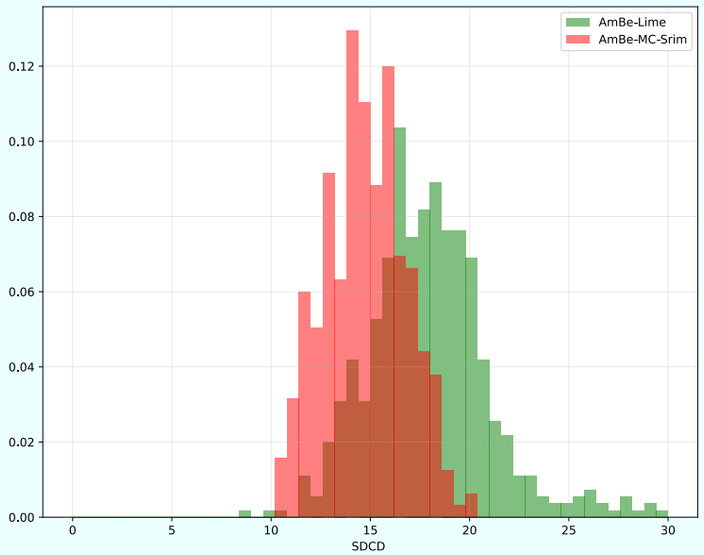


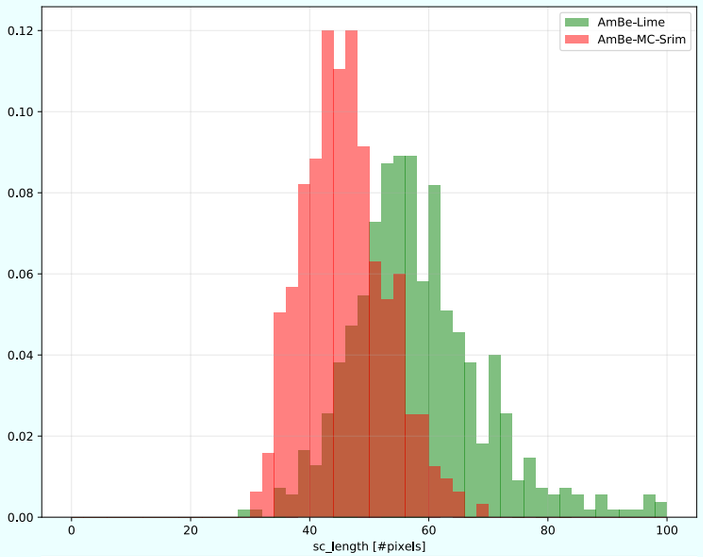


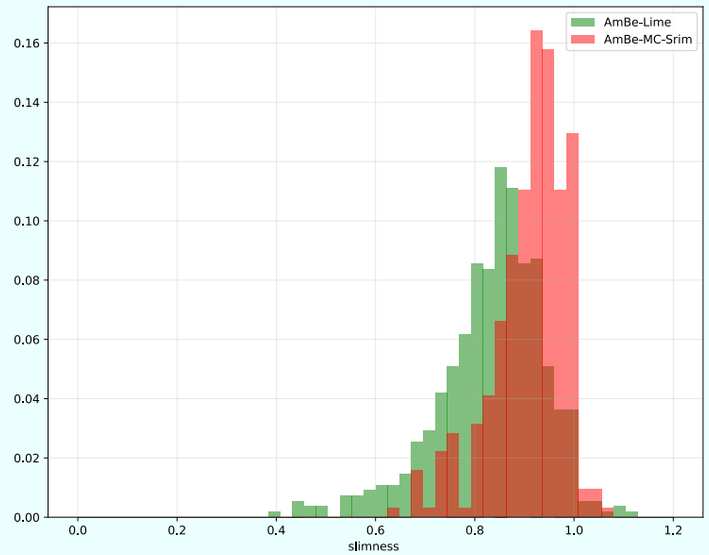


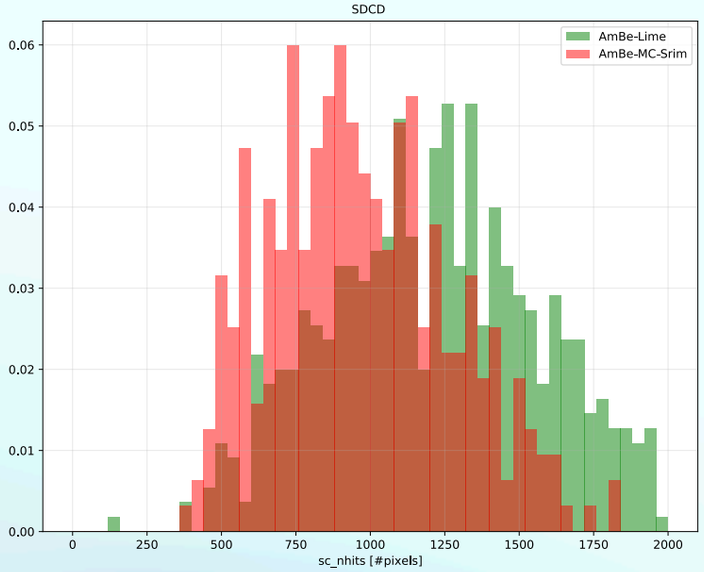
Events are selected with the same cut from both data and MC.

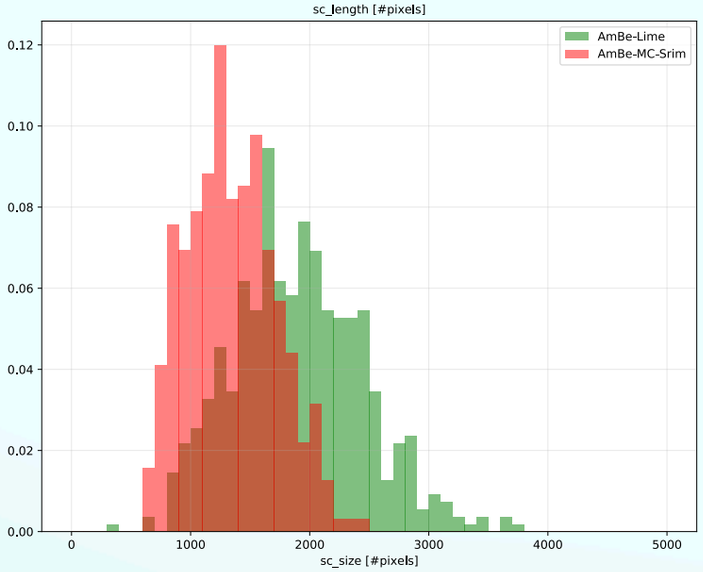
Shape variables comparison-NR

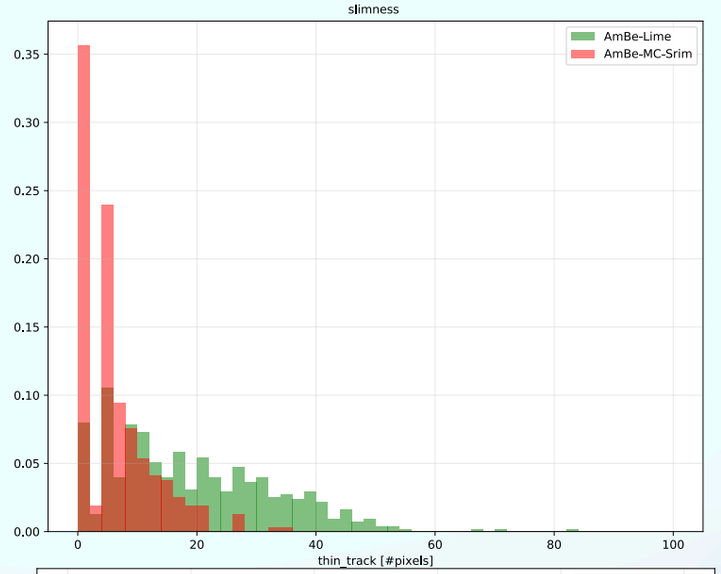


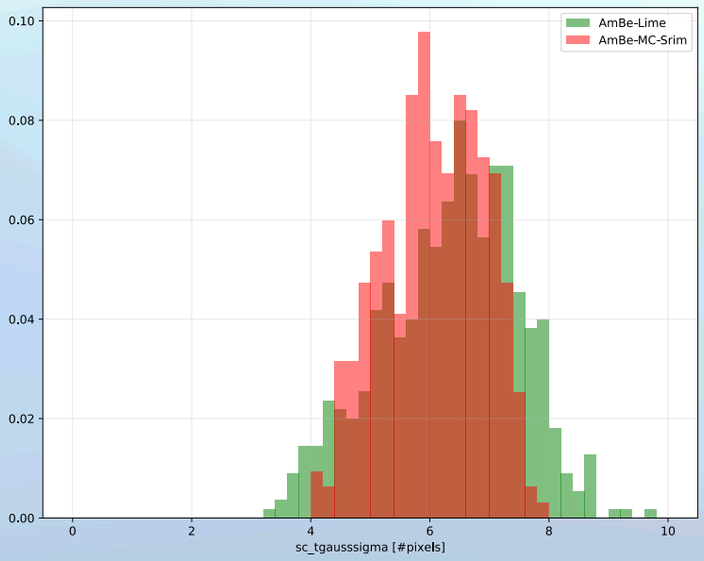


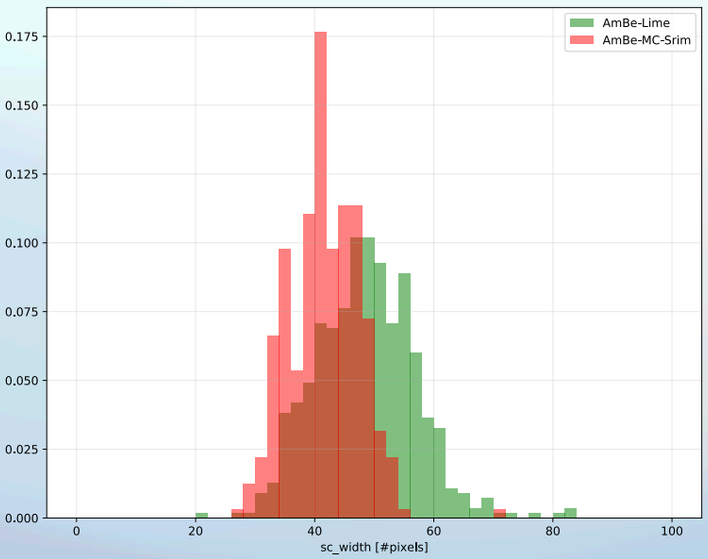


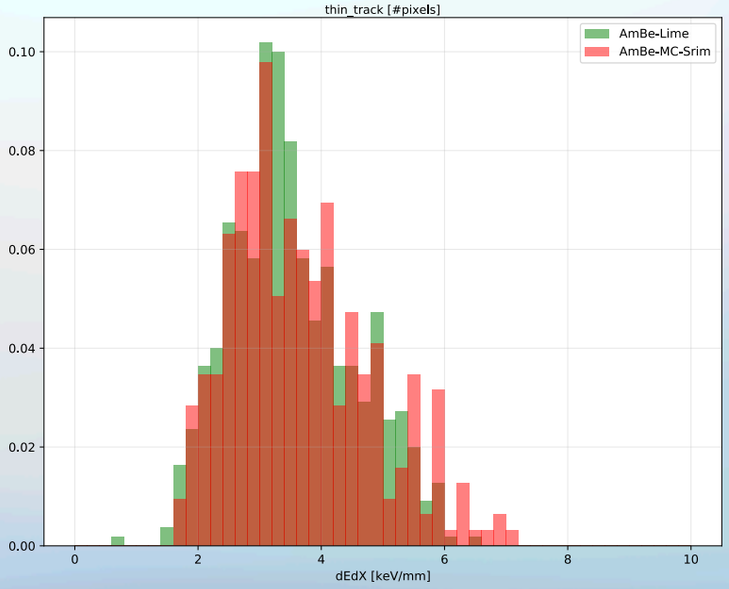




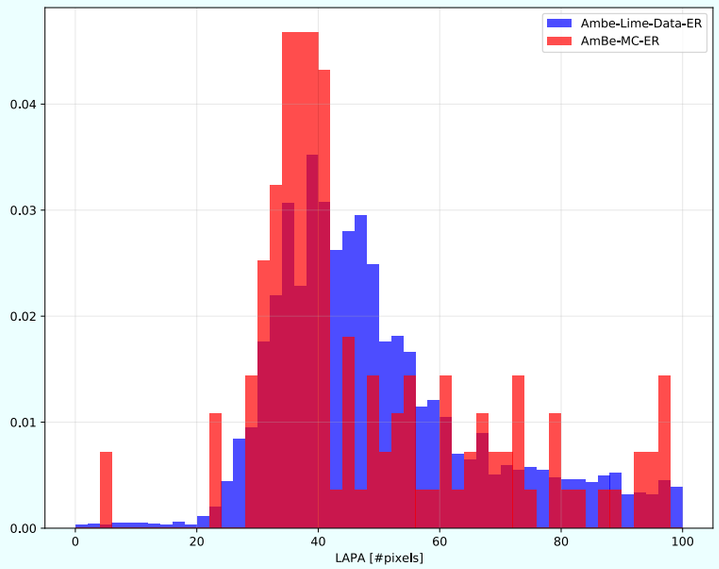


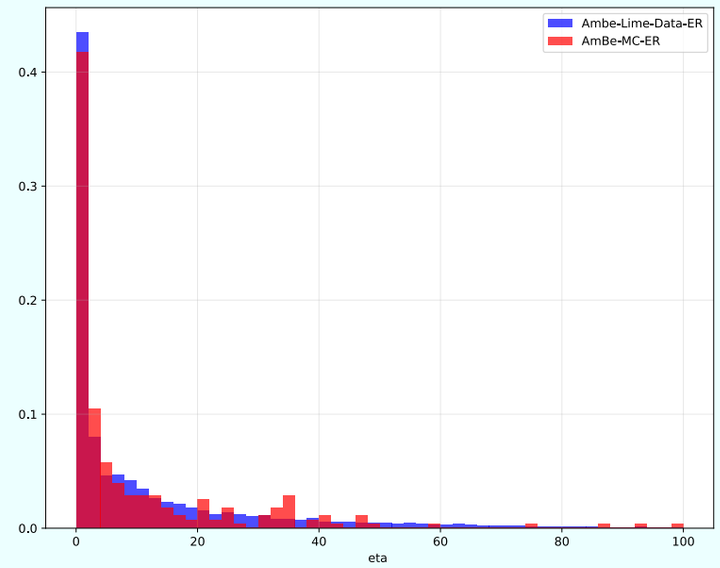


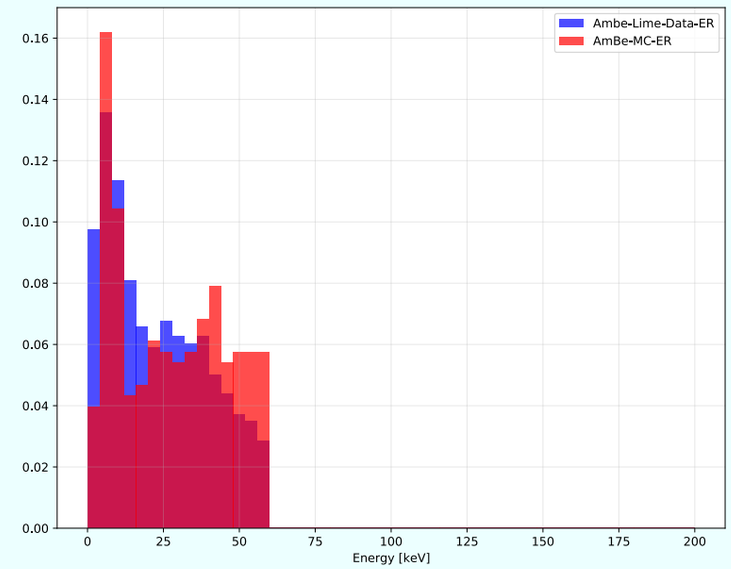


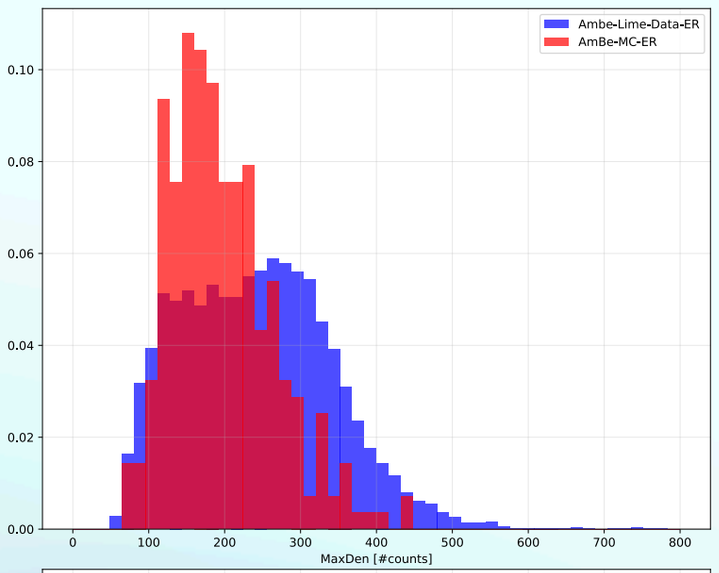


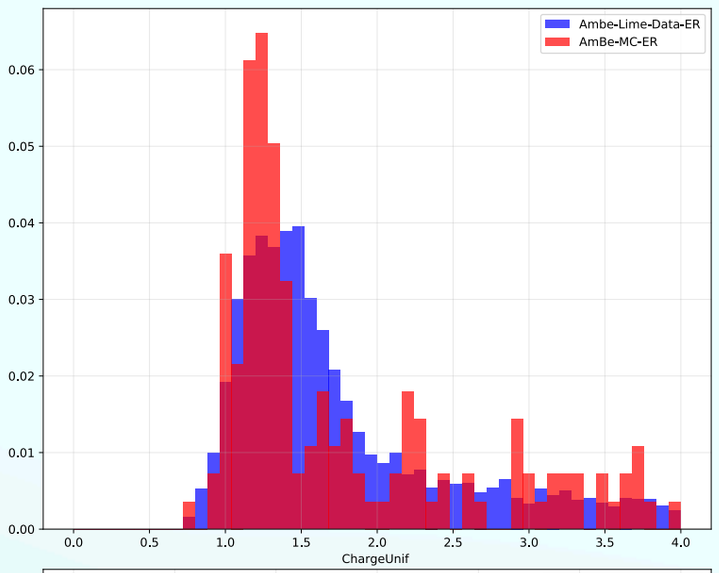
Shape variables comparison-ER

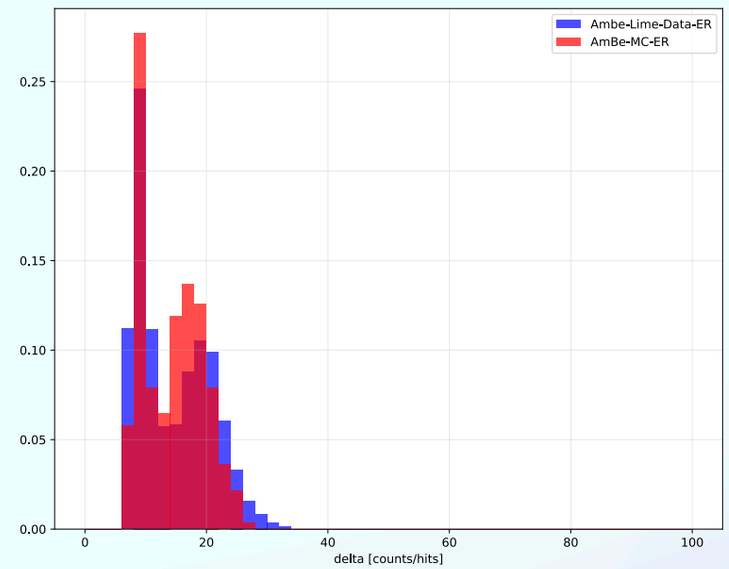


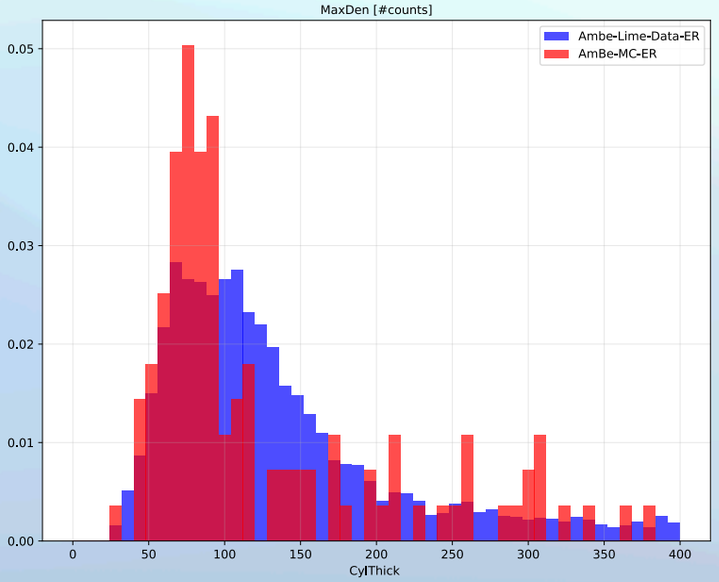


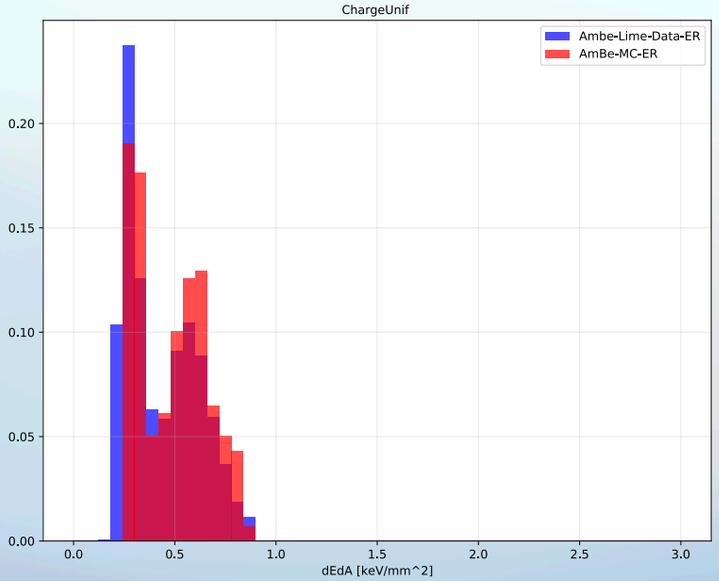






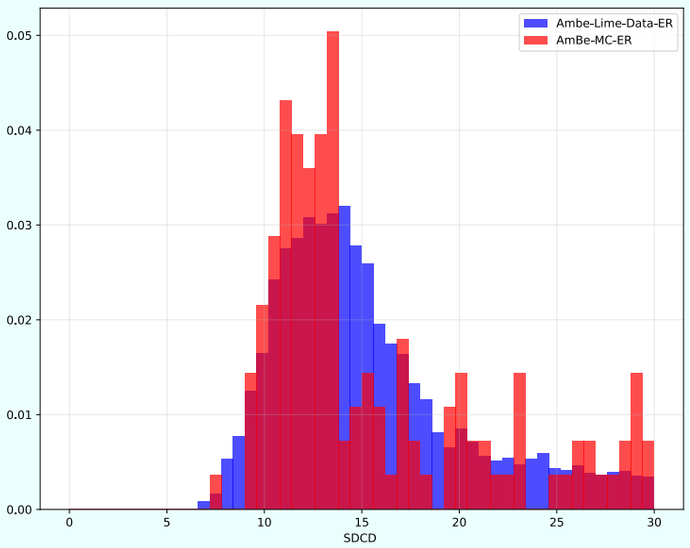


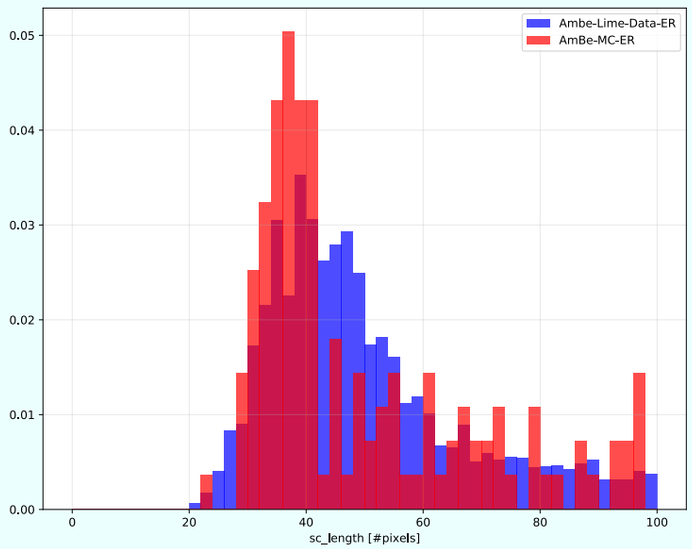


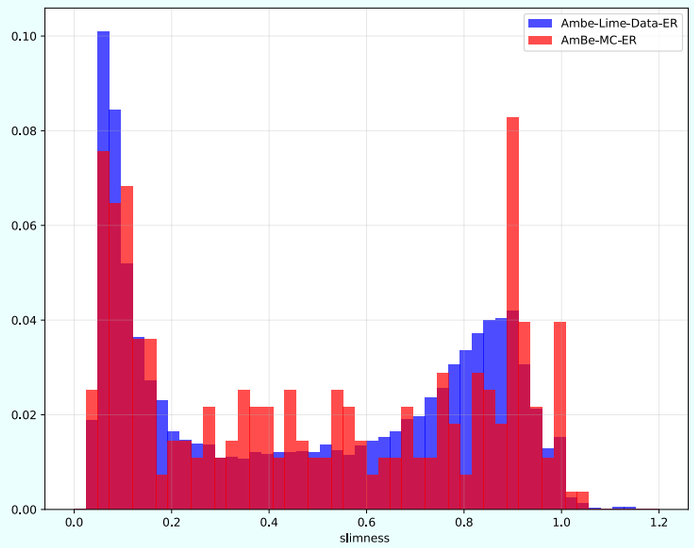


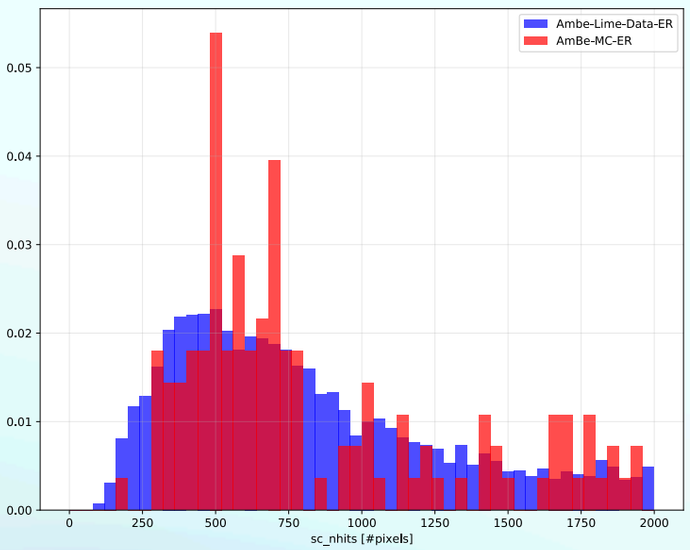
Events are selected with the same cut from both data and MC.

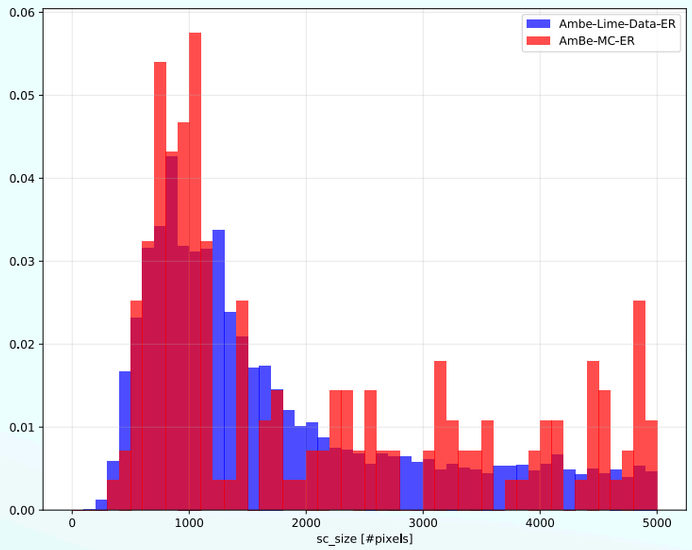
Shape variables comparison-ER

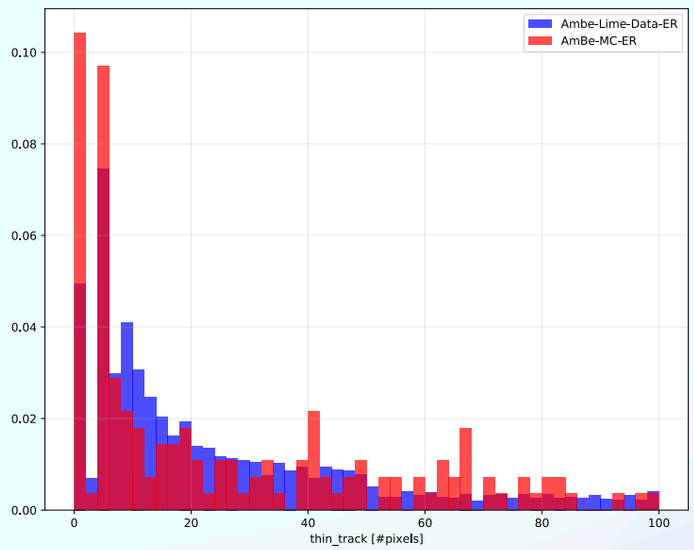


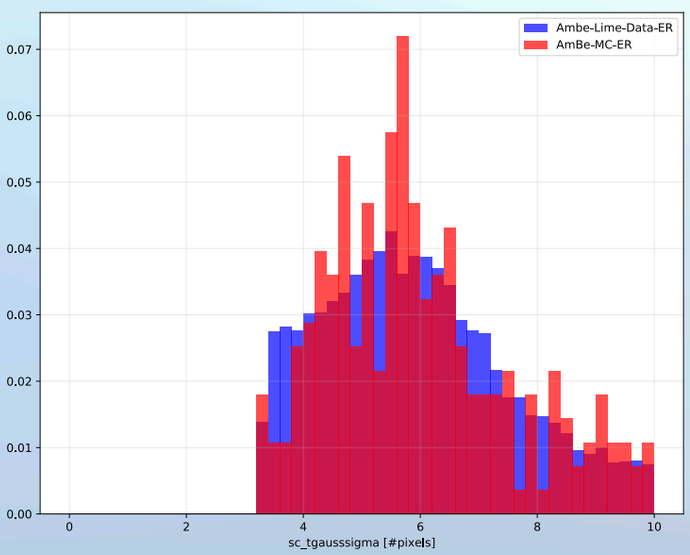


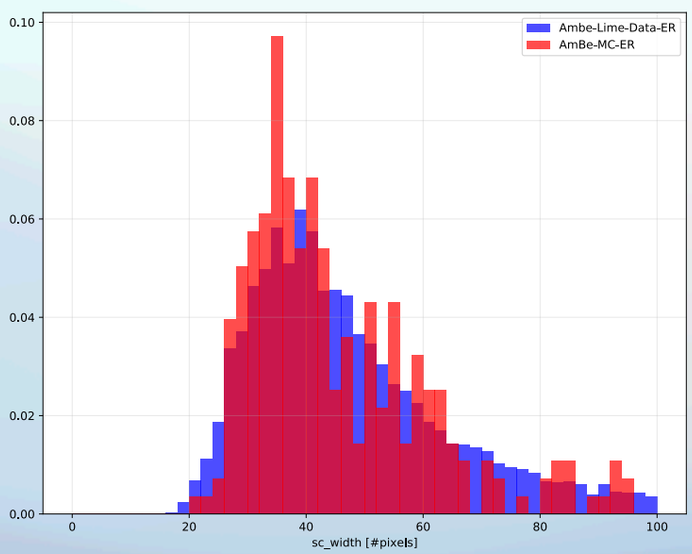


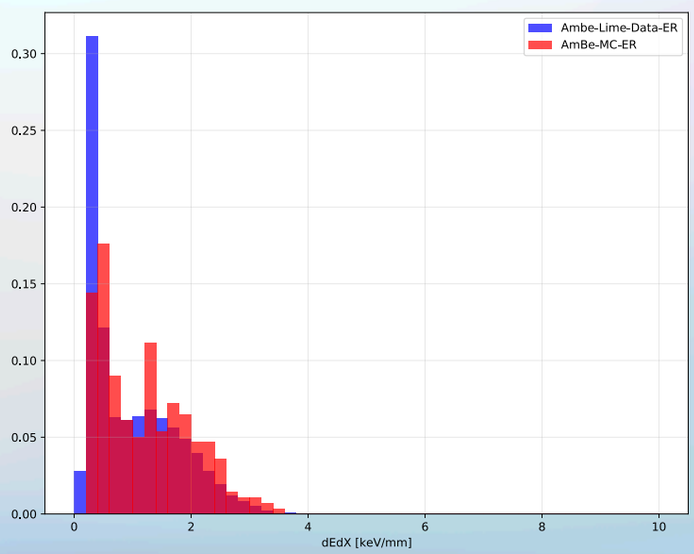








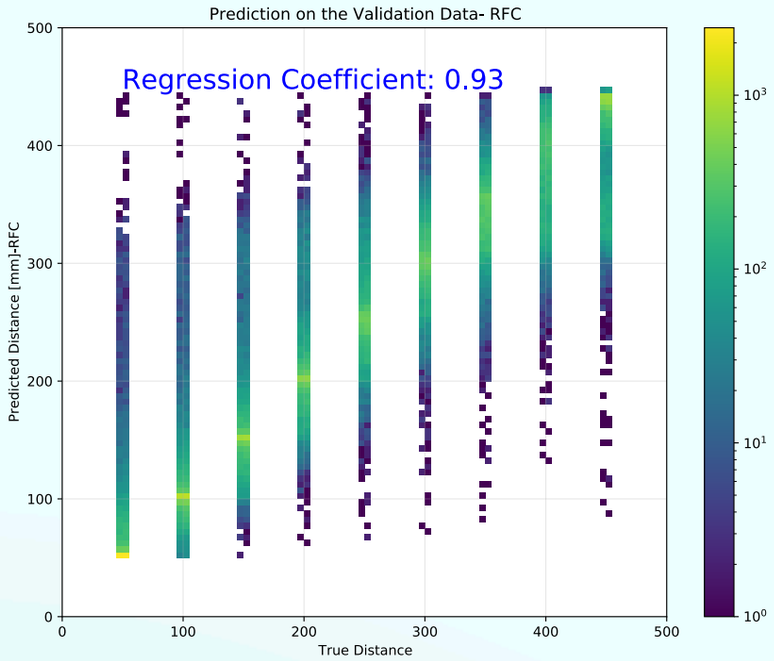


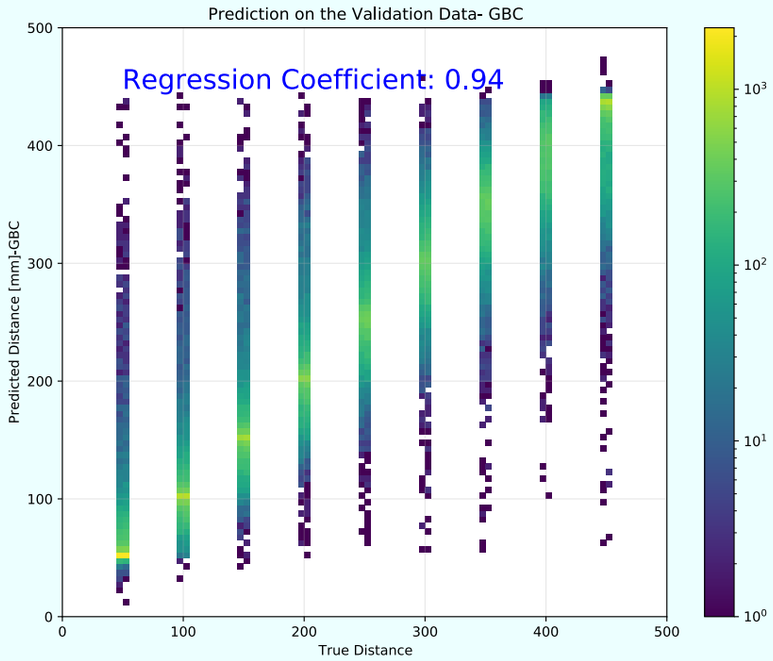


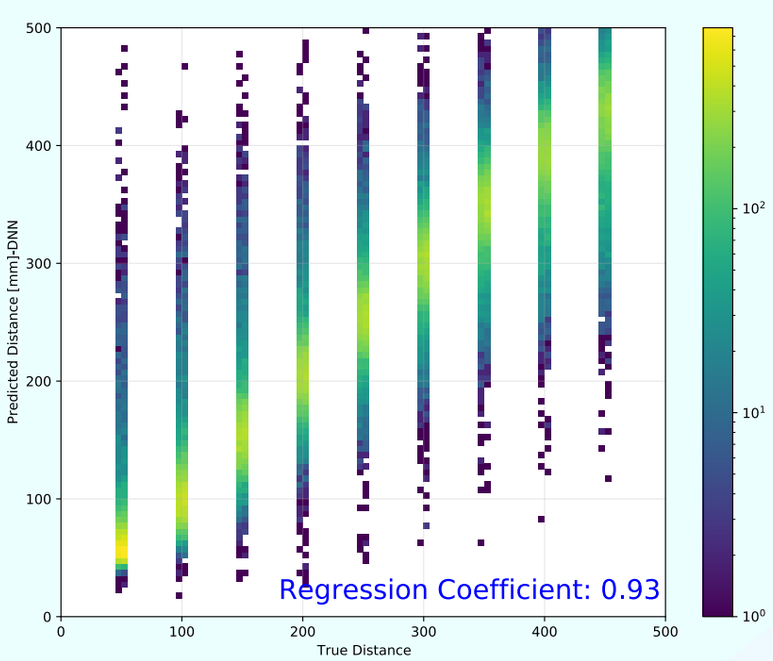
Z estimation using ML

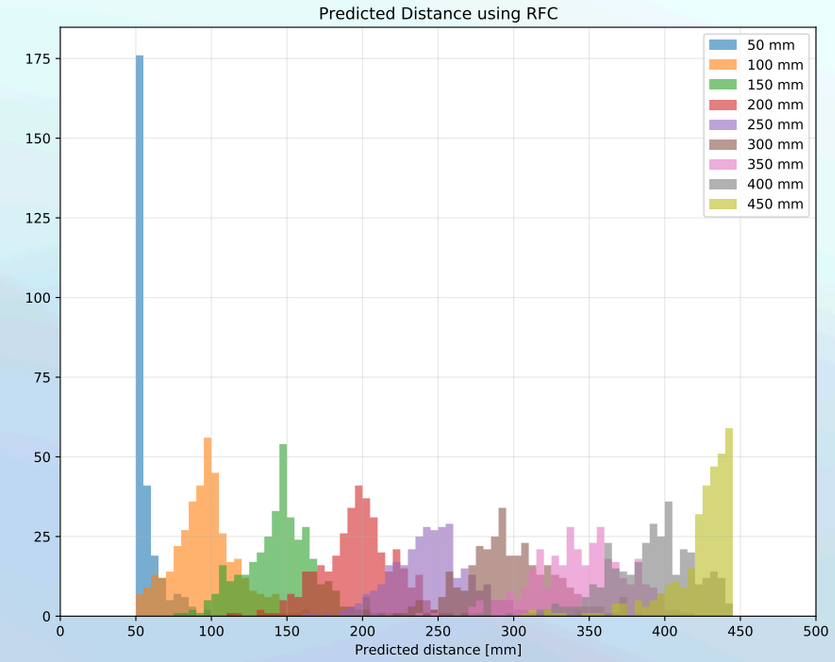
Training ML models for z prediction

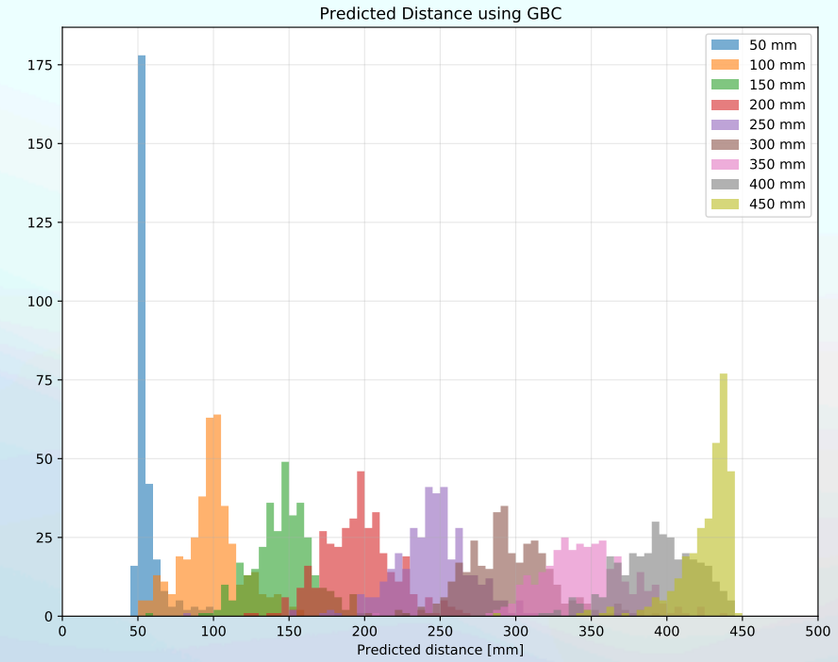
Variables = Energy, tgausssigma, length, width, size, nhits

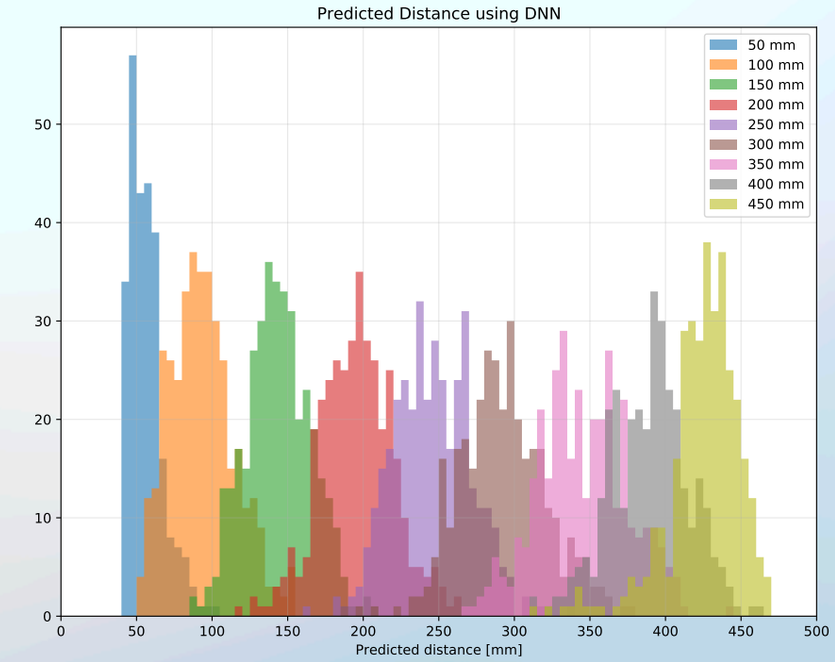




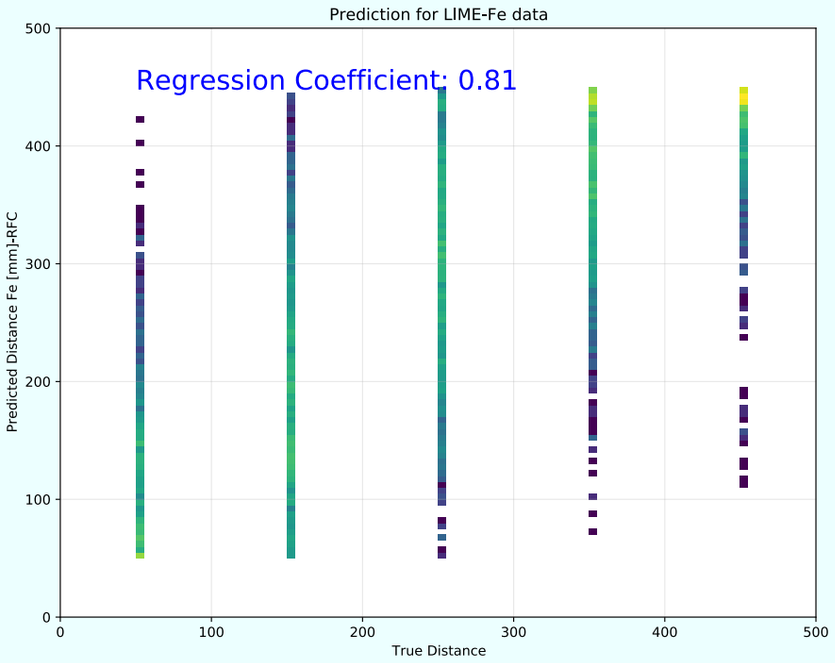


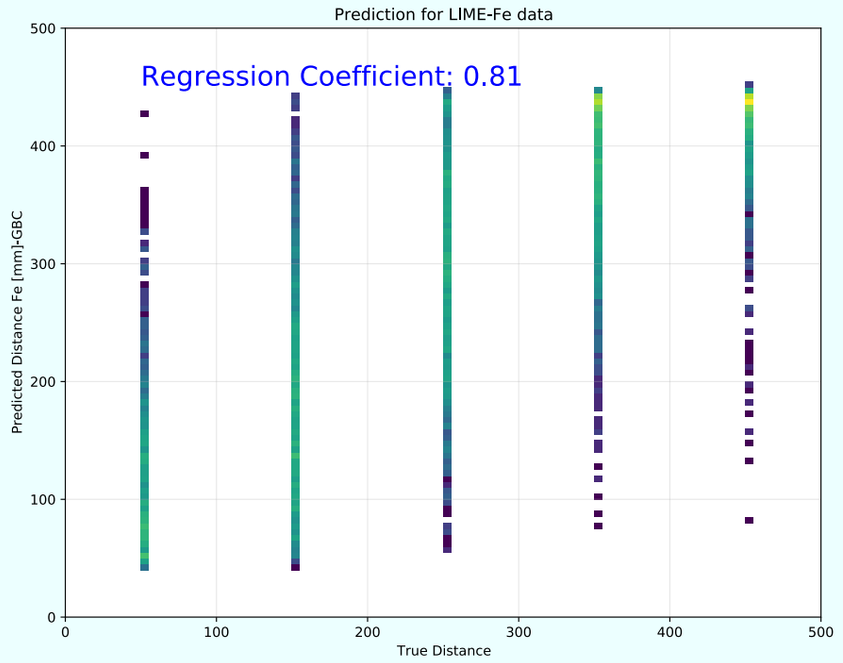


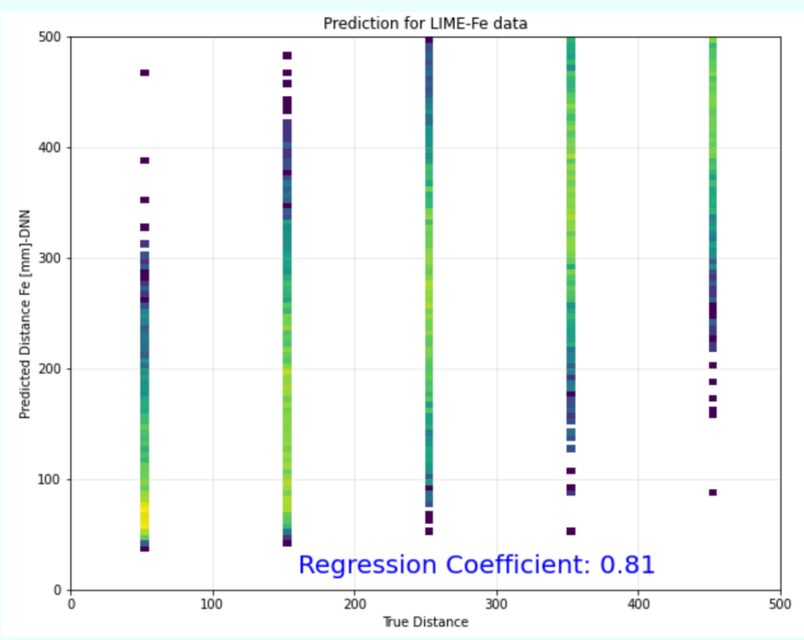


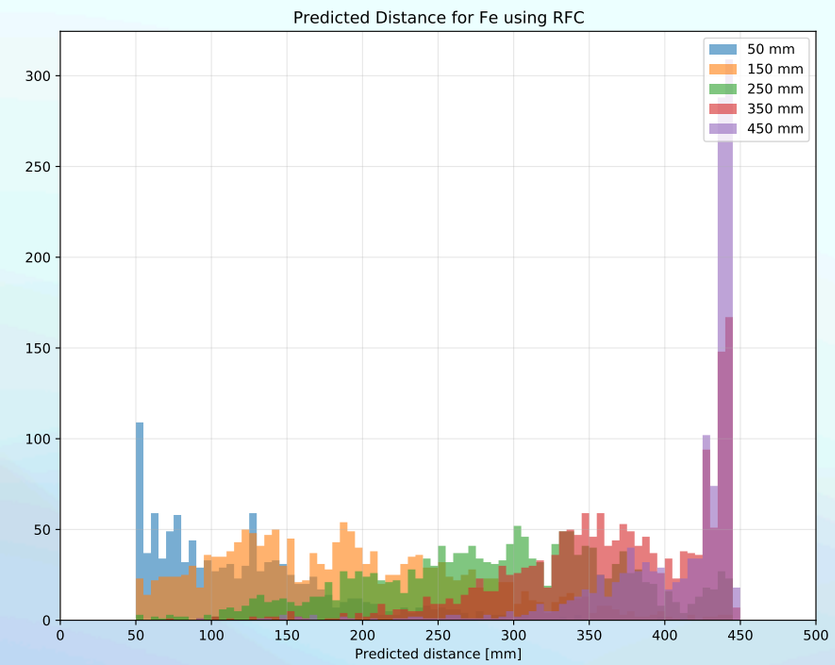


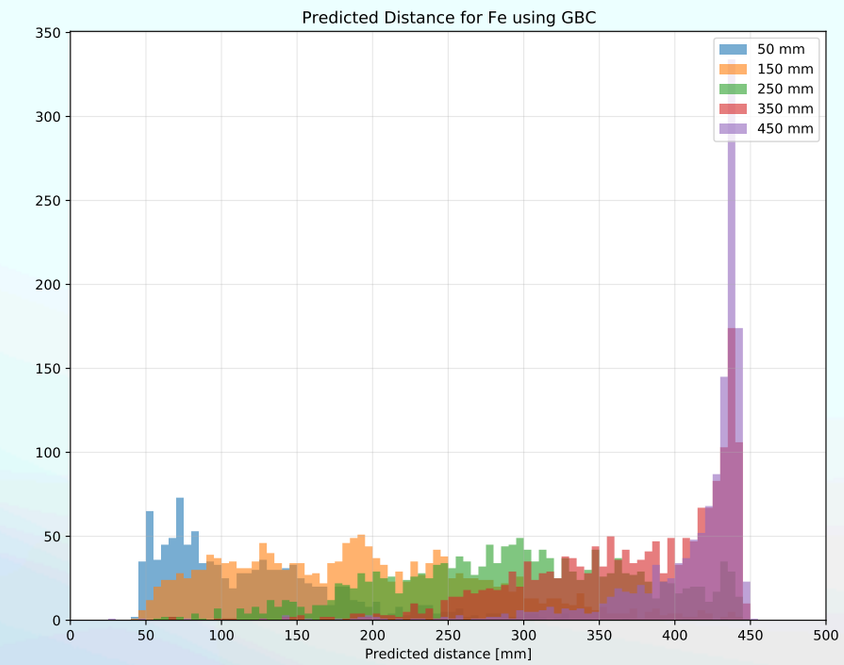
Testing ML models on Fe data

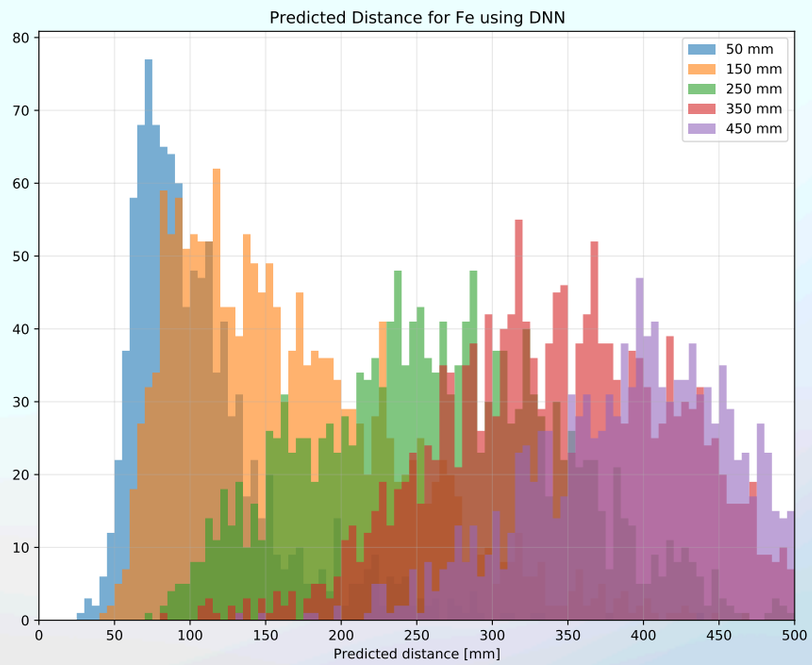












- Prediction on MC is better because the z in MC does not have the spread like the Fe data.