Comparison Metrics

Jan Scharf

EPICAL-2 Meeting, 11.07.2024

Overview: ML energy reconstruction

Idea: Increase complexity of data and model for better resolution.

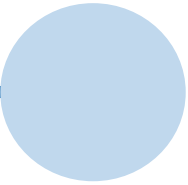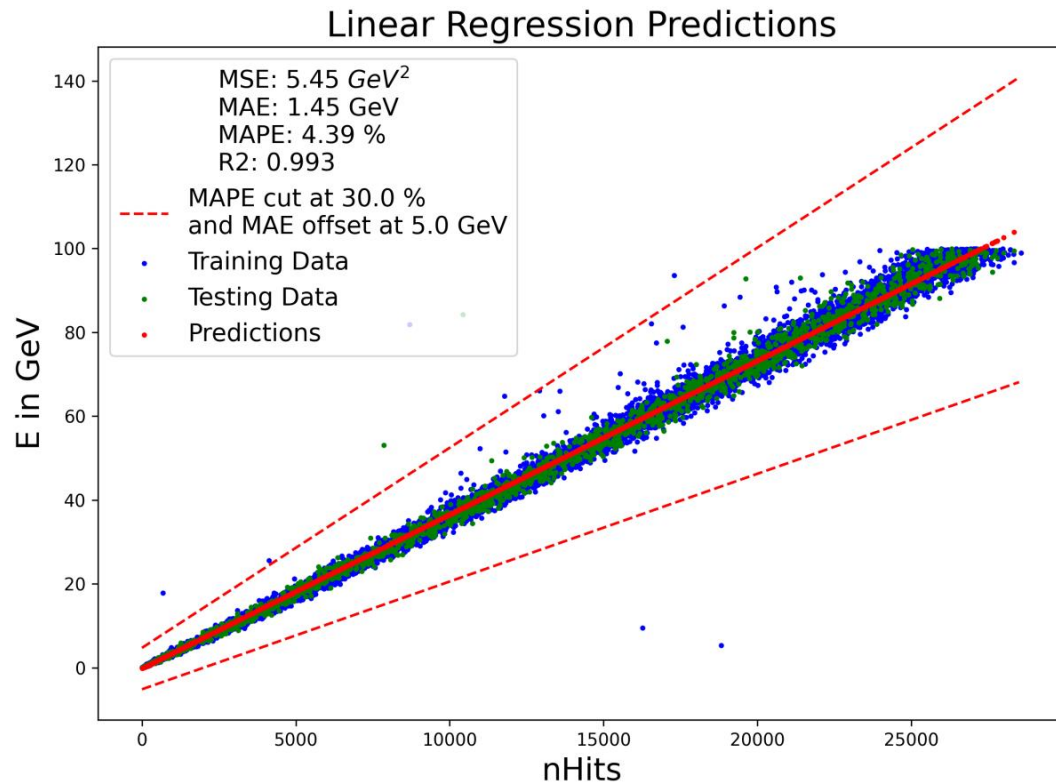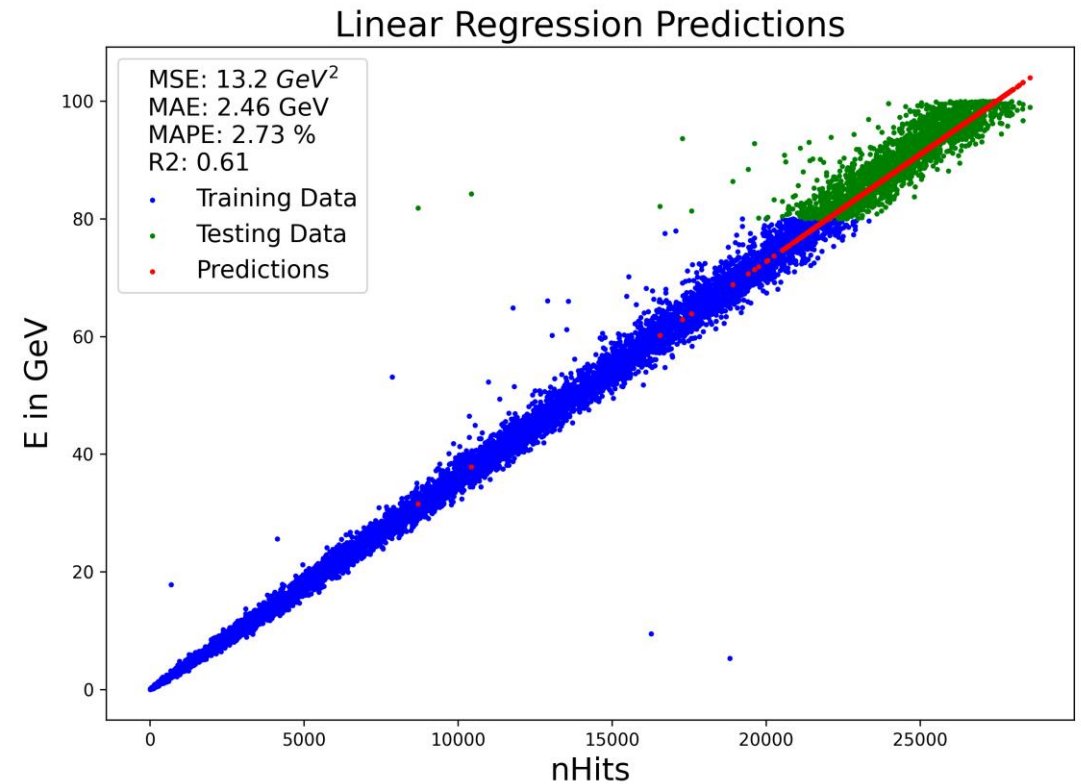|	
0-dim	• Linear regression on nHits • Chose metrics • Dataset exploration
1-dim	• Fit of longitudinal distribution • Complex implemenation strategies
3-dim	• Fit on full 3d hit distribution • Data preparation: cylindrical coordinate

Linear Regression Model

Training on all energies

Training on 0-80 GeV energies

Performance Metrics 1

- Mean Squared Error: $MSE = \frac{1}{N}\sum_{i=0}^{N-1}(y_{true,i} - y_{pred,i})^2$

  - Can also be written as the variance (sensitivity of the model to changes in the dataset) and bias (complexity of the model) of the model

- Mean Absolute Error: $MAE = \frac{1}{N}\sum_{i=0}^{N-1}|y_{true,i} - y_{pred,i}|$
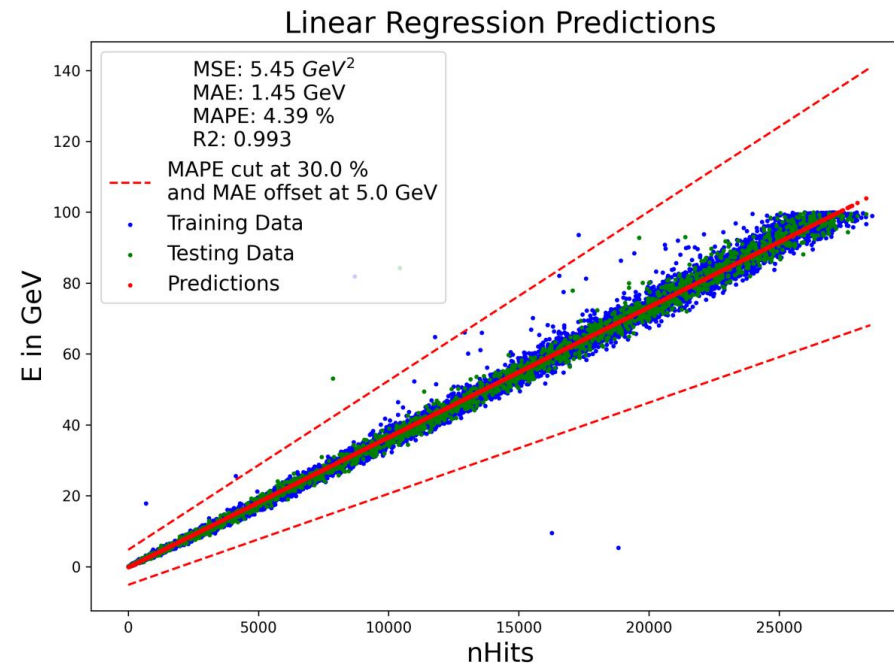
- Mean Absolute Percentage Error:

$$MAPE = \frac{1}{N}\sum_{i=0}^{N-1}\left|\frac{y_{true,i} - y_{pred,i}}{y_{true,i}}\right|$$
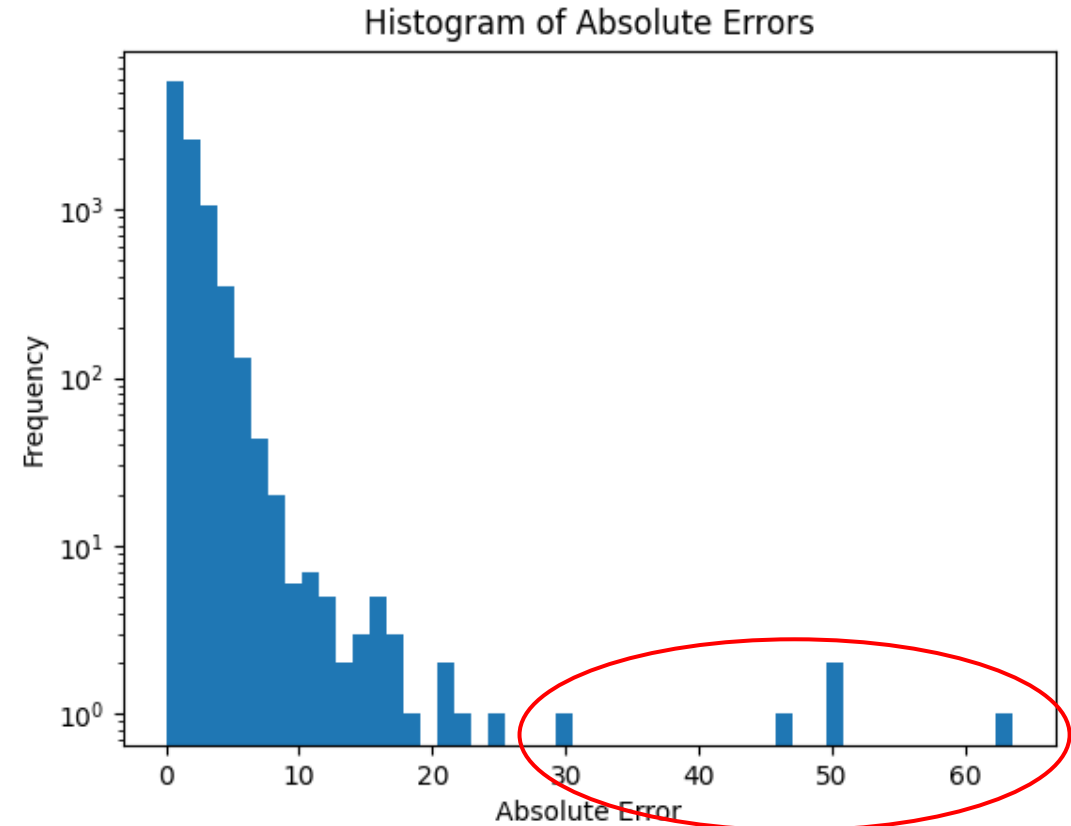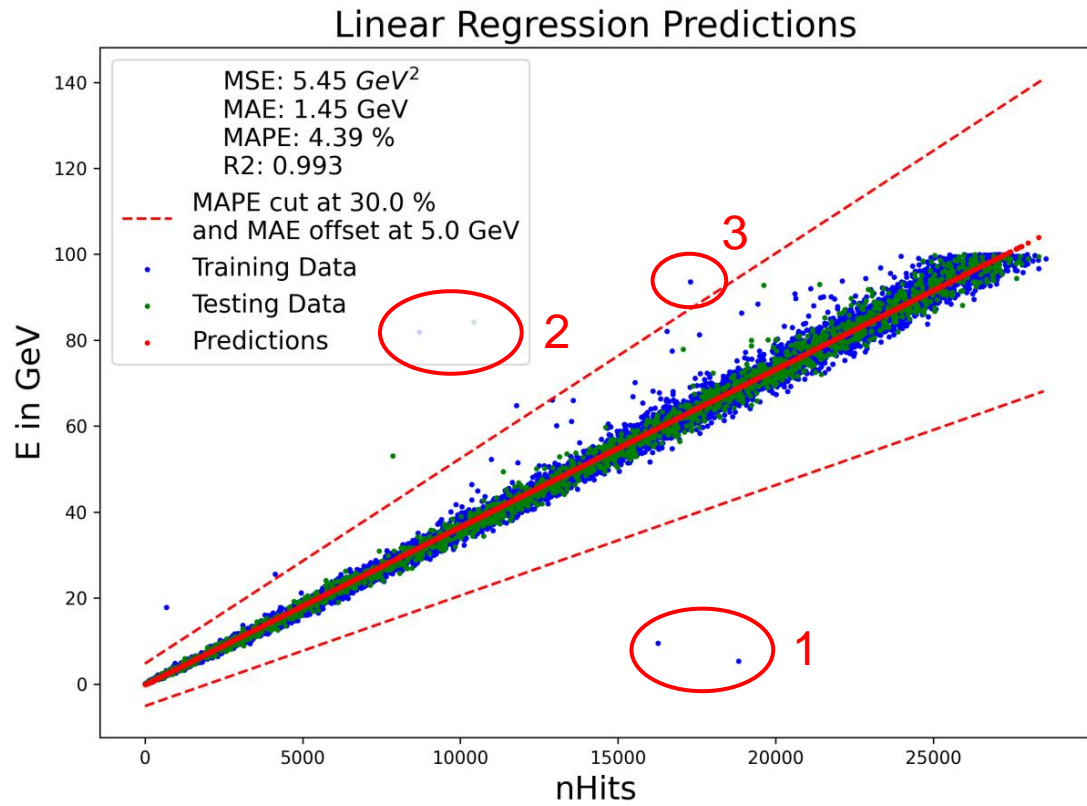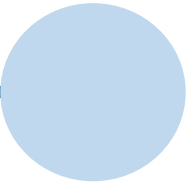
Performance Metrics 2

- R2-Score: $R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$

  - Measures the dependence of the regression on the input (here nHits)
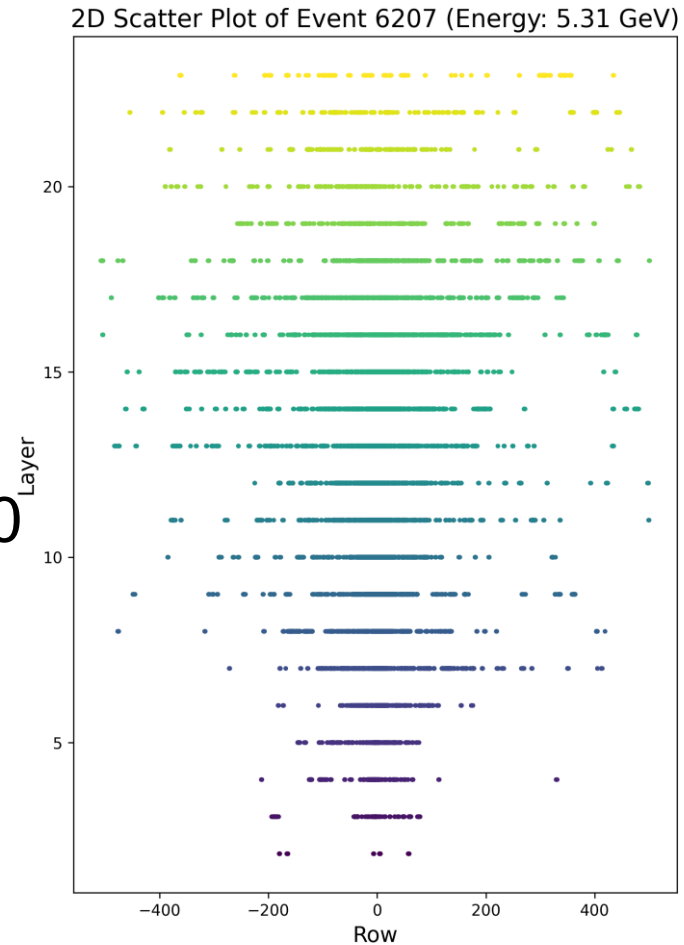 - $R^2 = 1$ is the best, $R^2 = 0$ only predicts the mean value of the dataset
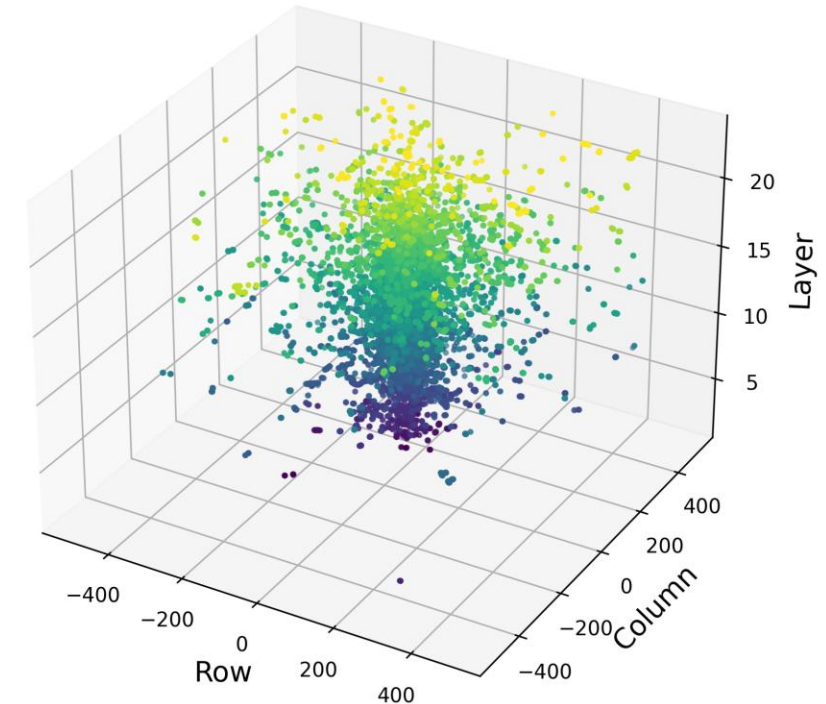 - Comparison of variance
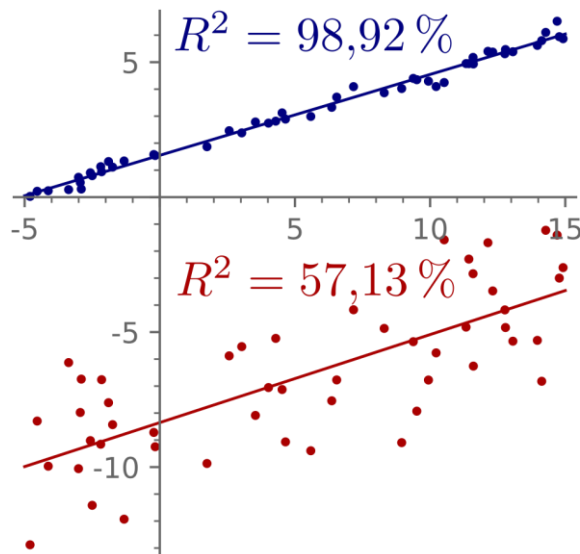
Outlier Events

Event 6207

- No hits in layer 0 and 1
- Simulation tags energy of secondary particle
- Energy is not right
- Other event only has a backscatter particle in layer 0

2D Scatter Plot of Event 6207 (Energy: 5.31 GeV)

3D Scatter Plot of Event 6207 (Energy: 5.31 GeV)

Events 7474 & 8320

2D Scatter Plot of Event 7474 (Energy: 81.8 GeV)

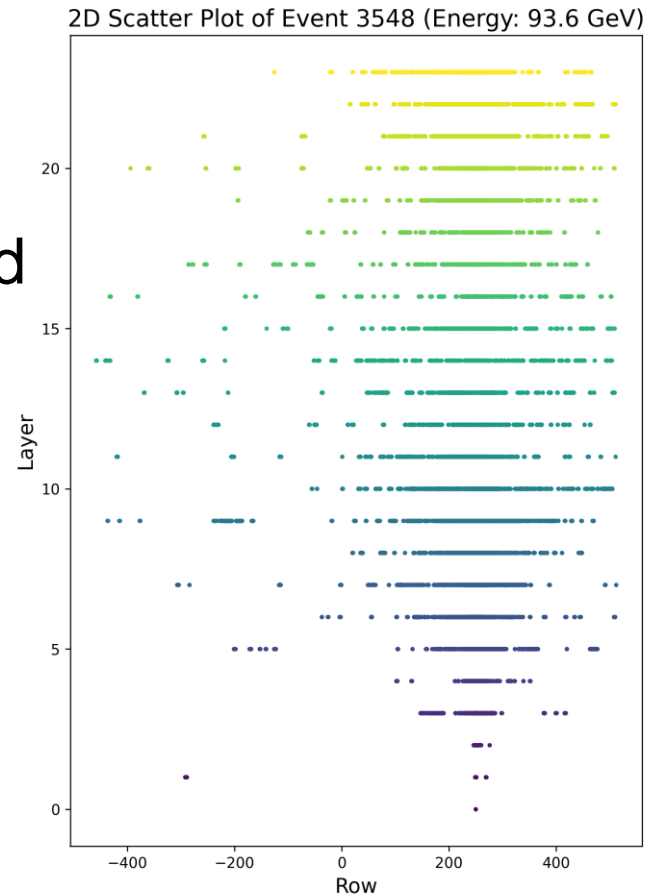2D Scatter Plot of Event 8320 (Energy: 84.2 GeV)
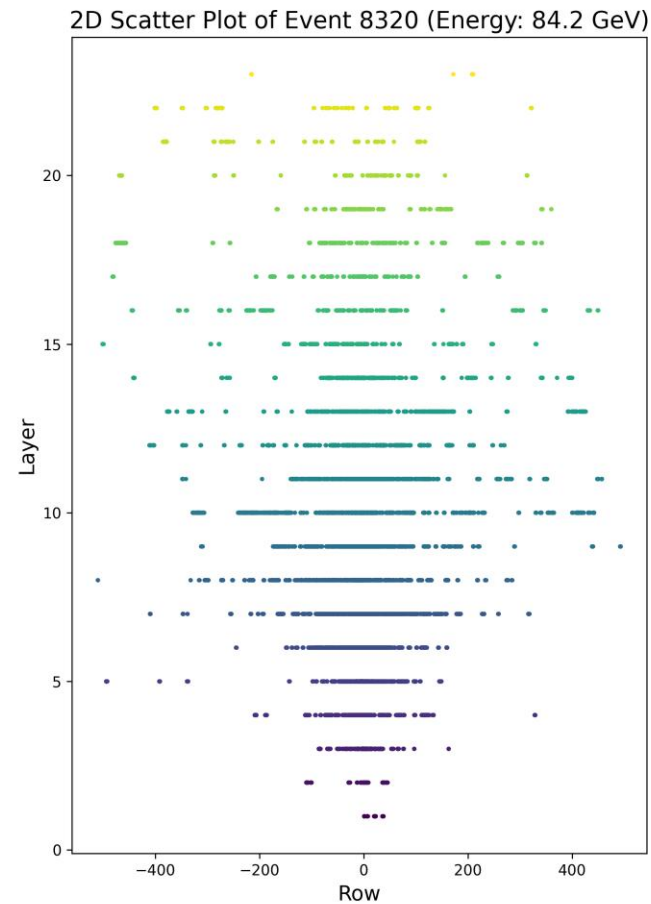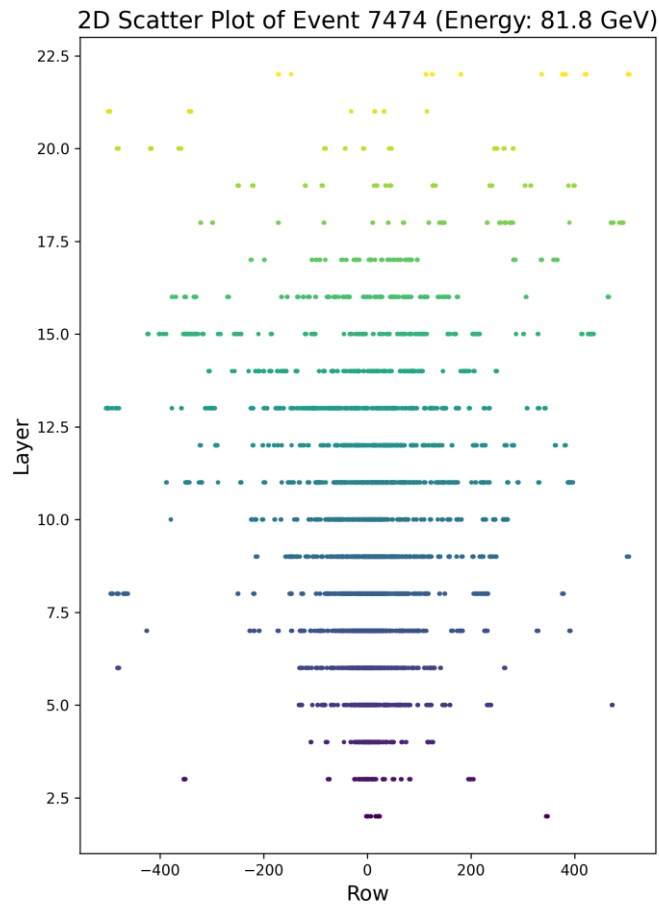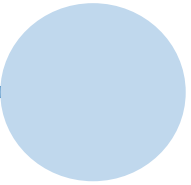
- Too few hits

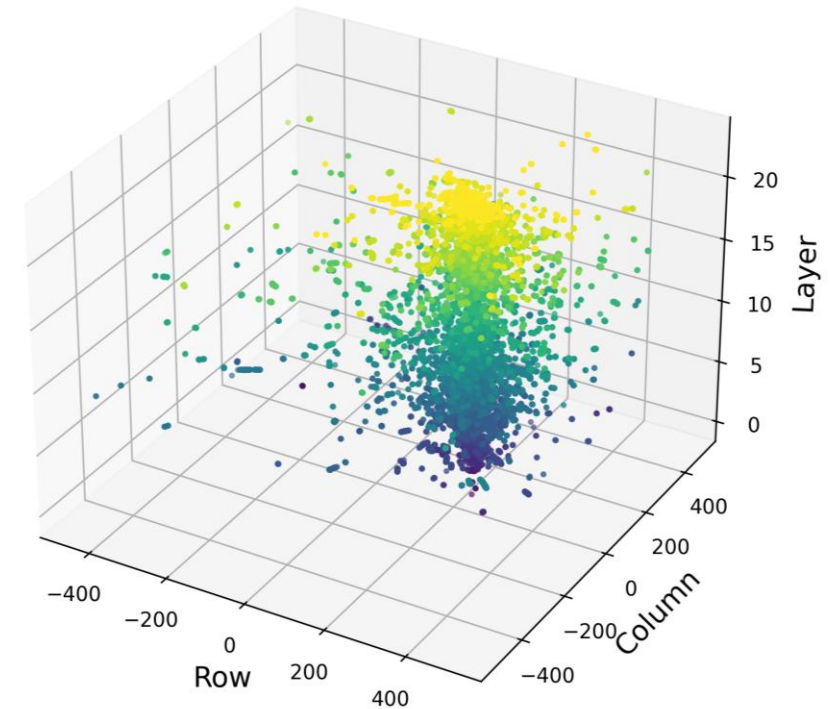- No hit in layer 0

- Problem ?

Event 3548

- Too few hits

- Shower is long, but not broad

- Problem: Particle not completely showered in detector

2D Scatter Plot of Event 3548 (Energy: 93.6 GeV)

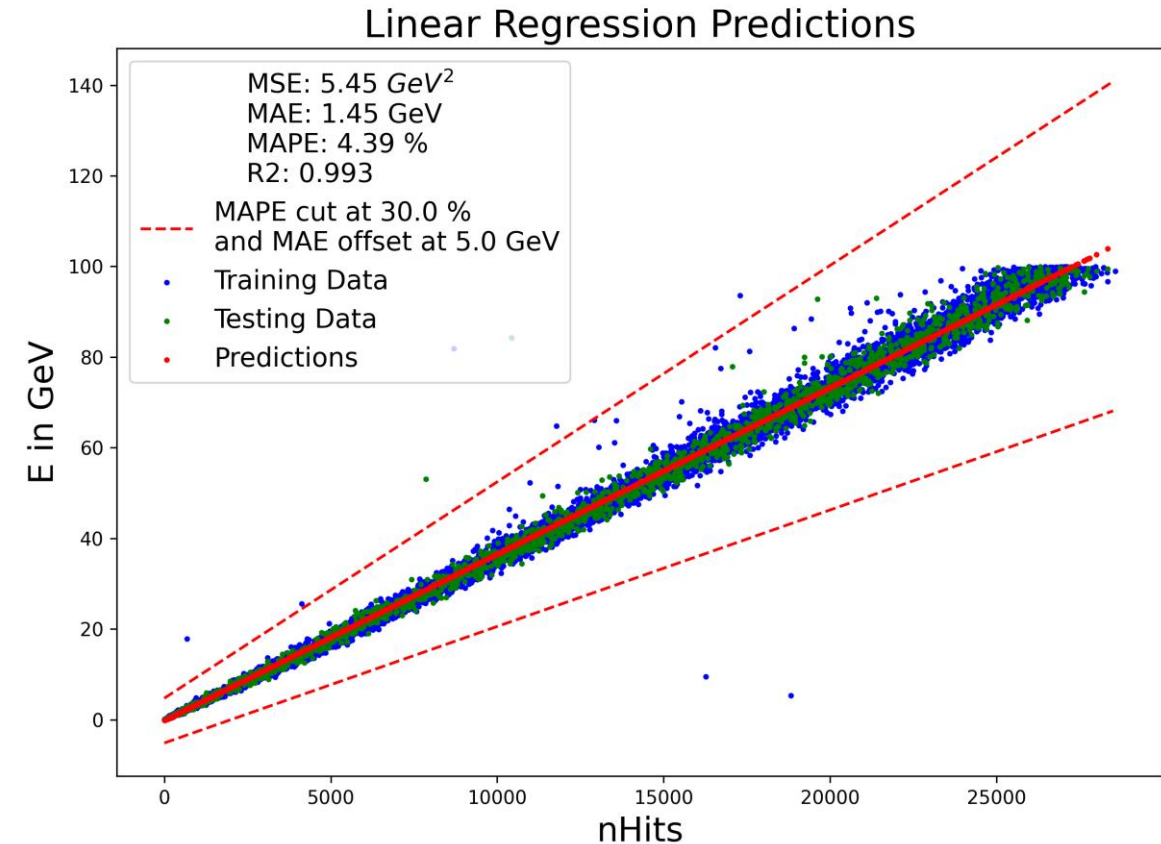3D Scatter Plot of Event 3548 (Energy: 93.6 GeV)

Exclude Outlier Events

- Dataset should exclude these events for future models

- Cut based on linear regression model

- Only data with energy of defined distance to predictions, E in range:

$$[\hat{E} + b * \hat{E} + \mathrm{a}, \qquad \hat{E} - b * \hat{E} - \mathrm{a}]$$

- b = 30%, a = 5 GeV

?

Linear Regression Predictions

MSE: 5.45 GeV^2
MAE: 1.45 GeV
MAPE: 4.39 %
R2: 0.993
MAPE cut at 30.0 % and MAE offset at 5.0 GeV
Training Data
Testing Data
Predictions

EPICAL-2 logo ???