

Machine learning techniques for \overline{He} research in cosmic rays

Francesco Rossi

Introduction

- Large *He* background
- \overline{He} are rare events, as a rule of thumb $1 : 10^9 He$
- No signal model available, only 4He Monte Carlo and ISS-data.

Goals:

- 1) Study *He* background and find charge confusion sources
- 2) Develop tools to reduce *He* contributions.

Goals:

- 1) Study *He* background and find charge confusion sources
- 2) Develop tools to reduce *He* contributions.

Strategy :

G1)

- Use *He* BC1236 Monte Carlo samples (L1-focused and L1-L9 focused)
- Select “well reconstructed” ${}^4\text{He}$ events and study the charged confused events

G2)

- Train a classifier (on MC) to recognise different charge confusion sources
- Search for “outliers” in data using an anomaly detection technique.
- Combine the machine learning techniques

Study *He* background and find charge
confusion sources.

Monte carlo selection (He B1236 L1-focused and L1-L9 focused)

IsPhysicsTrigger

TRIGGER

β clusters ≥ 4

$\beta > 0$

$\chi_{COO}^2 < 4$

β

Track number ≥ 1

track pattern L1&L2&(L3|L4)&(L5|L6)&(L7|L8) (≥ 5)

TRACK

charge YJ (inner) $\in [1.7, 2.4]$

charge YJ (L1) $\in [1.6, 3.0]$

Inner fiducial volume

L1 fiducial volume

$\chi_Y^2 < 10$

TRK

charge UTOF $\in [1.5, 3.0]$

charge LTOF ≥ 1.5

TOF charge

Inner rigidities signs

RIGIDITY

Tracker fiducial volume cut:

L1: $|R| < 62\text{cm}$, $|Y| < 47\text{cm}$

L2: $|R| < 62\text{cm}$, $|Y| < 40\text{cm}$

L3: $|R| < 46\text{cm}$, $|Y| < 44\text{cm}$

L4: $|R| < 46\text{cm}$, $|Y| < 44\text{cm}$

L5: $|R| < 46\text{cm}$, $|Y| < 36\text{cm}$

L6: $|R| < 46\text{cm}$, $|Y| < 36\text{cm}$

L7: $|R| < 46\text{cm}$, $|Y| < 44\text{cm}$

L8: $|R| < 46\text{cm}$, $|Y| < 44\text{cm}$

UH = Rig. [UH-inner]

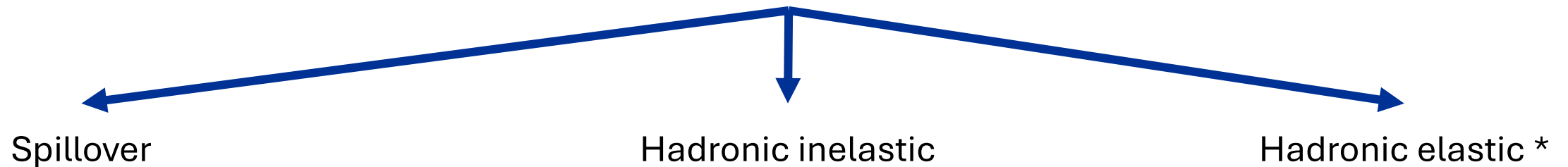
LH = Rig. [LH-inner]

Inner = Rig. [inner]

If($R_{inner} < 0$) \rightarrow (UH < 0 , LH < 0)

Sources of charged confused events

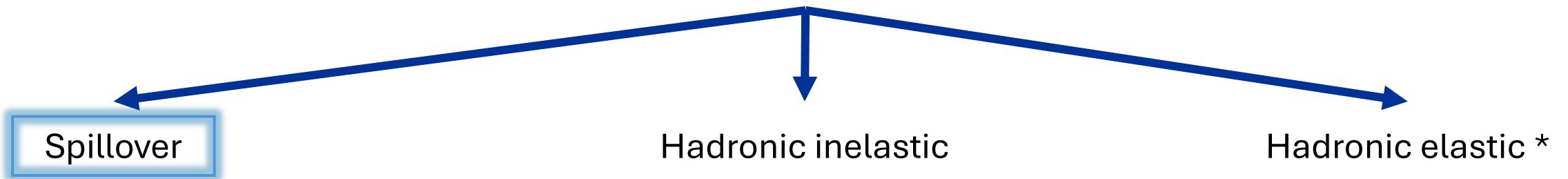
Using He Monte Carlo B1236 L1-focused and L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**



* Large angle scattering

Sources of charged confused events

Using He Monte Carlo B1236 L1-focused and L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**

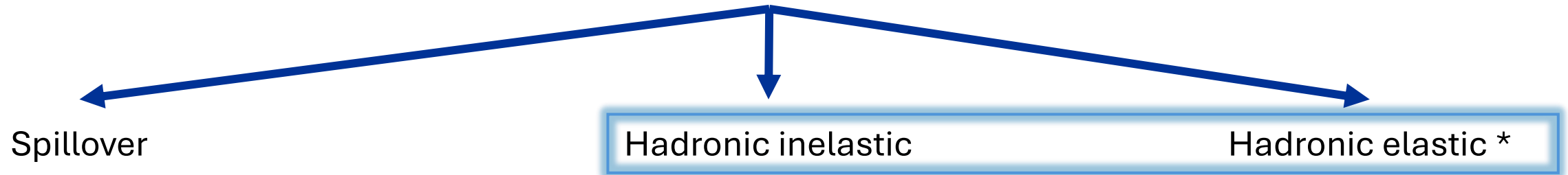


Silicon tracker finite resolution

* Large angle scattering

Sources of charged confused events

Using He Monte Carlo B1236 L1-focused and L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**

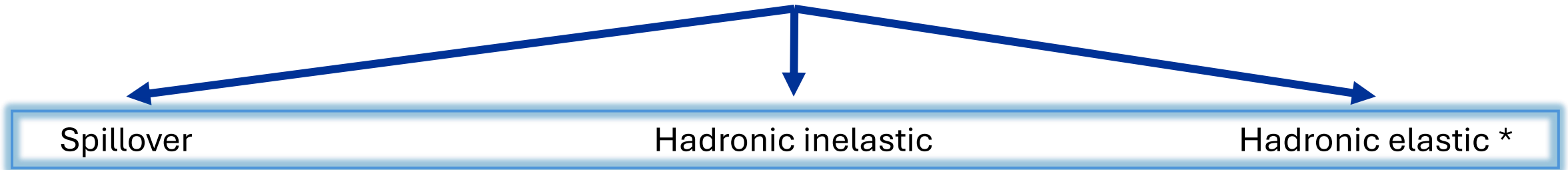


Interactions within the detector

* Large angle scattering

Sources of charged confused events

Using He Monte Carlo B1236 L1-focused and L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**

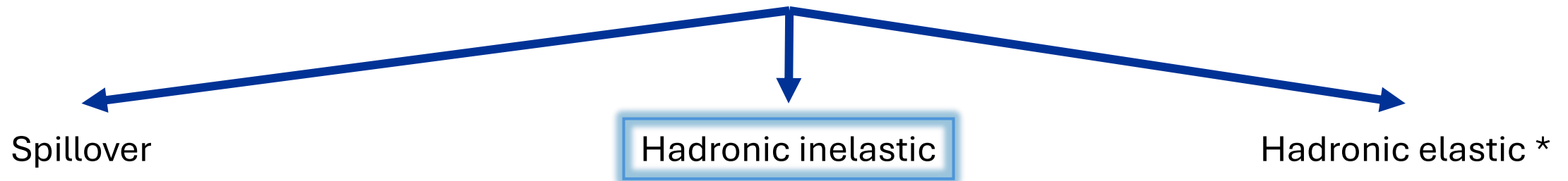


For each source, we select a sample to be used in the training of a classifier

* Large angle scattering

Sources of charged confused events

Using He Monte Carlo B1236 L1-focused and B1236 L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**

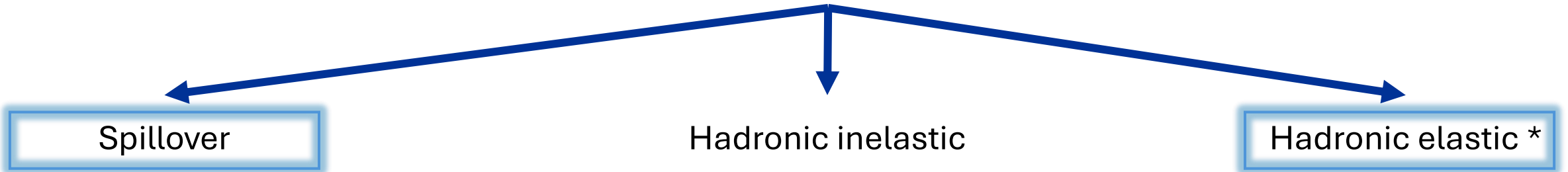


Search in the secondaries list looking for inelastic interaction products inside the inner tracker

* Large angle scattering

Sources of charged confused events

Using He Monte Carlo B1236 L1-focused and B1236 L1-L9 focused, and selecting the **reconstructed events with $R < 0$** , we identified **three sources of charge confusion**



Search in the secondaries list looking for secondaries produced inside the inner tracker (HasSecondary)

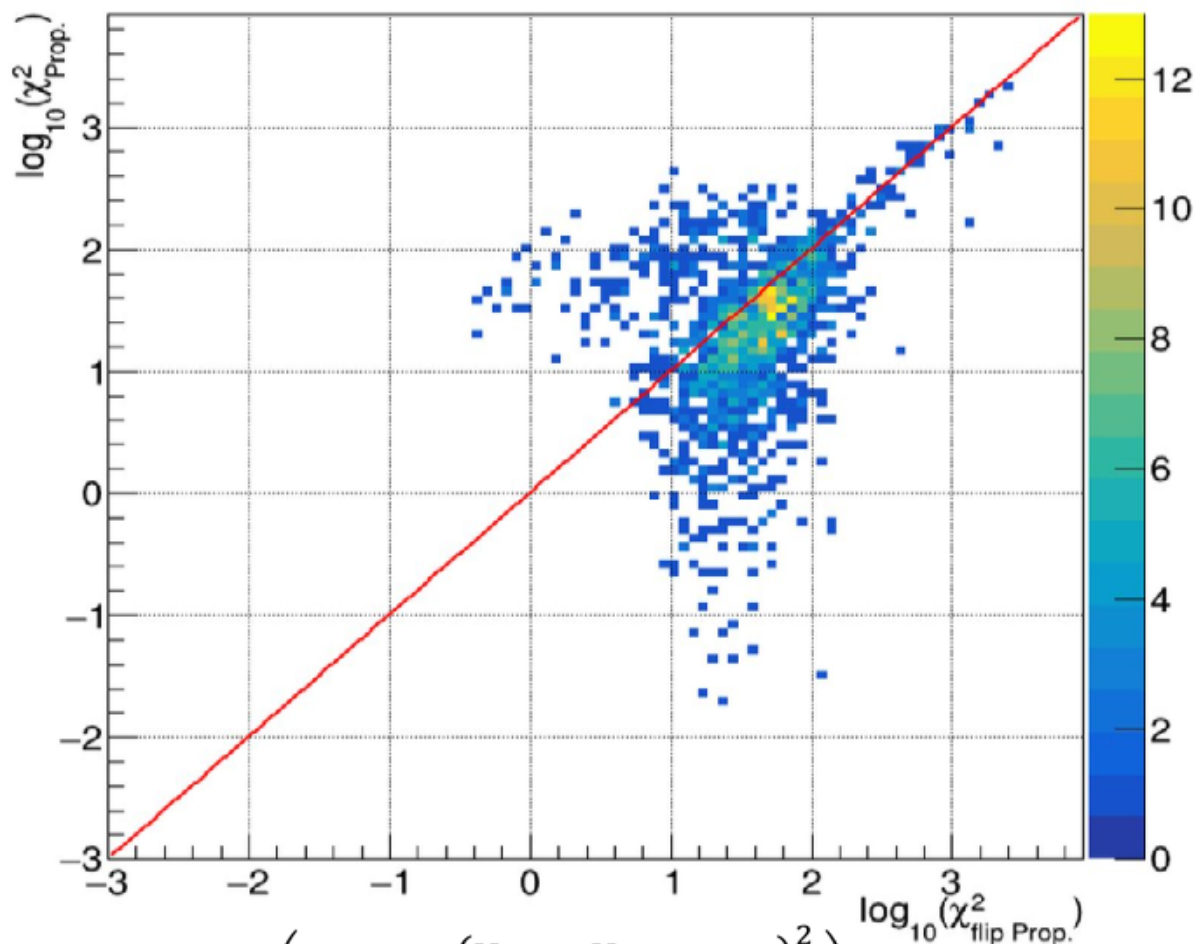
Events are not Had. Inel. Interactions and primary nuclei reaches L2

Propagation of two tracks: $R_{true}(L2)$ and $R_{inner} (< 0)$
Build two χ^2 comparing y coordinate with MC true info on each layer.

$\frac{\chi_{R_{true}}^2}{\chi_{R_{inner}}^2} \geq 1.05$	→ El. scat.
$0.95 \geq \frac{\chi_{R_{true}}^2}{\chi_{R_{inner}}^2}$	→ (HasSecondary) ? Other : Spillover
$0.95 < \frac{\chi_{R_{true}}^2}{\chi_{R_{inner}}^2} < 1.05$	→ (HasSecondary) ? Other : Spillover

$R_{inner} < 0$ (no secondaries within inner tracker)

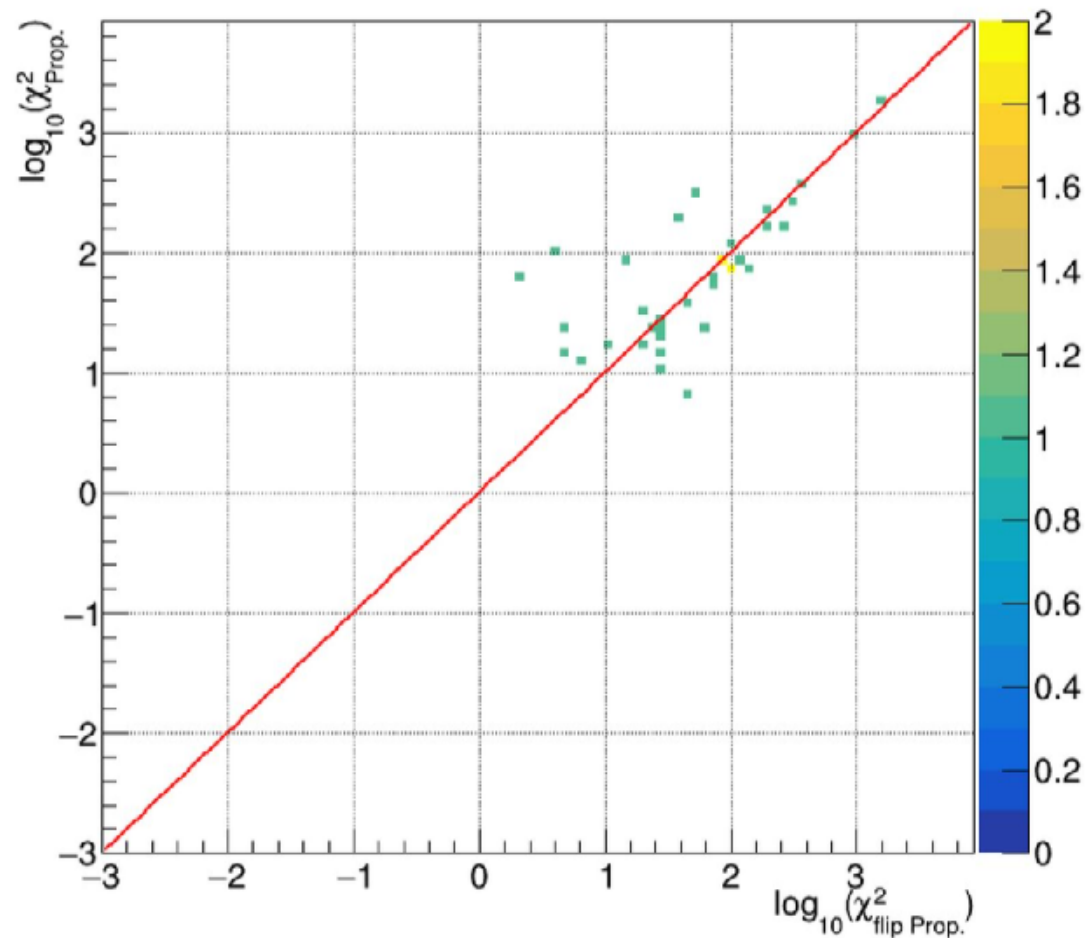
N_FlipProp_bin_(104.337868_130.184189)_xy



$$\chi_{prop}^2 = \frac{\left(\sum_{i-layer} \frac{(Y_{MC} - Y_{prop.track})^2}{\sigma_{hit}^2} \right)}{Ndof}$$

$R_{inner} < 0$ (elastic scattering within inner tracker)

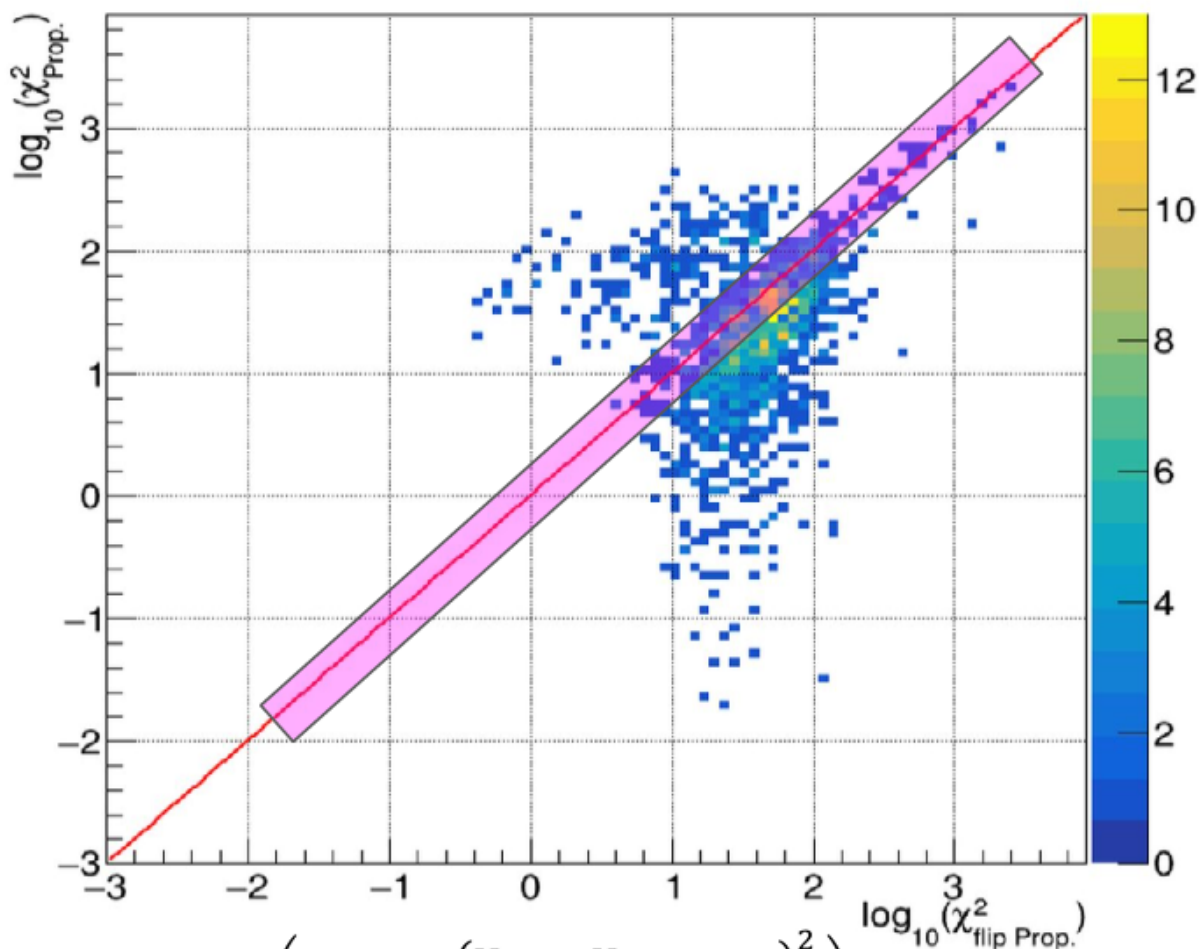
N_FlipProp_bin_(104.337868_130.184189)_xy



$$\sigma_{hit}^2 = 15 \mu m$$

$R_{\text{inner}} < 0$ (no secondaries within inner tracker)

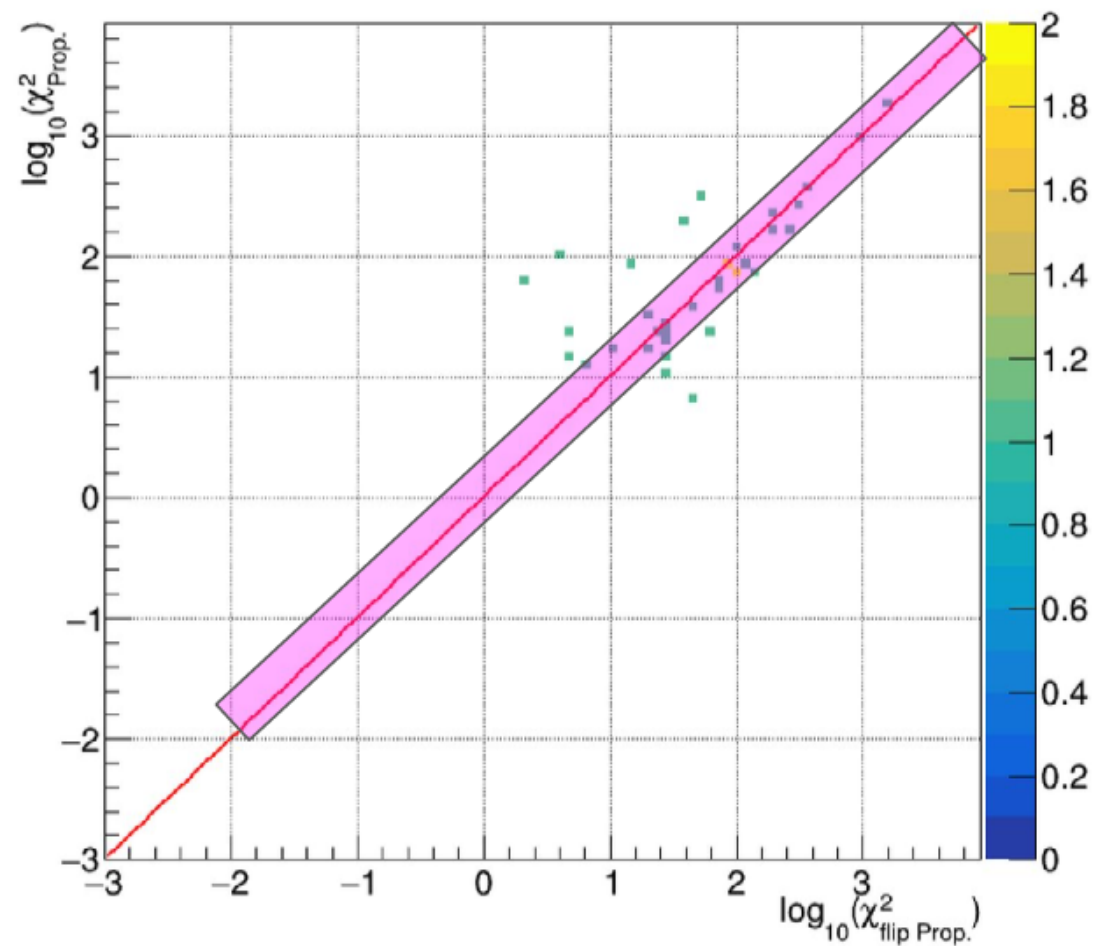
N_FlipProp_bin_(104.337868_130.184189)_xy



$$\chi_{\text{prop}}^2 = \frac{\left(\sum_{i\text{-layer}} \frac{(Y_{MC} - Y_{\text{prop.track}})^2}{\sigma_{\text{hit}}^2} \right)}{N_{\text{dof}}}$$

$R_{\text{inner}} < 0$ (elastic scattering within inner tracker)

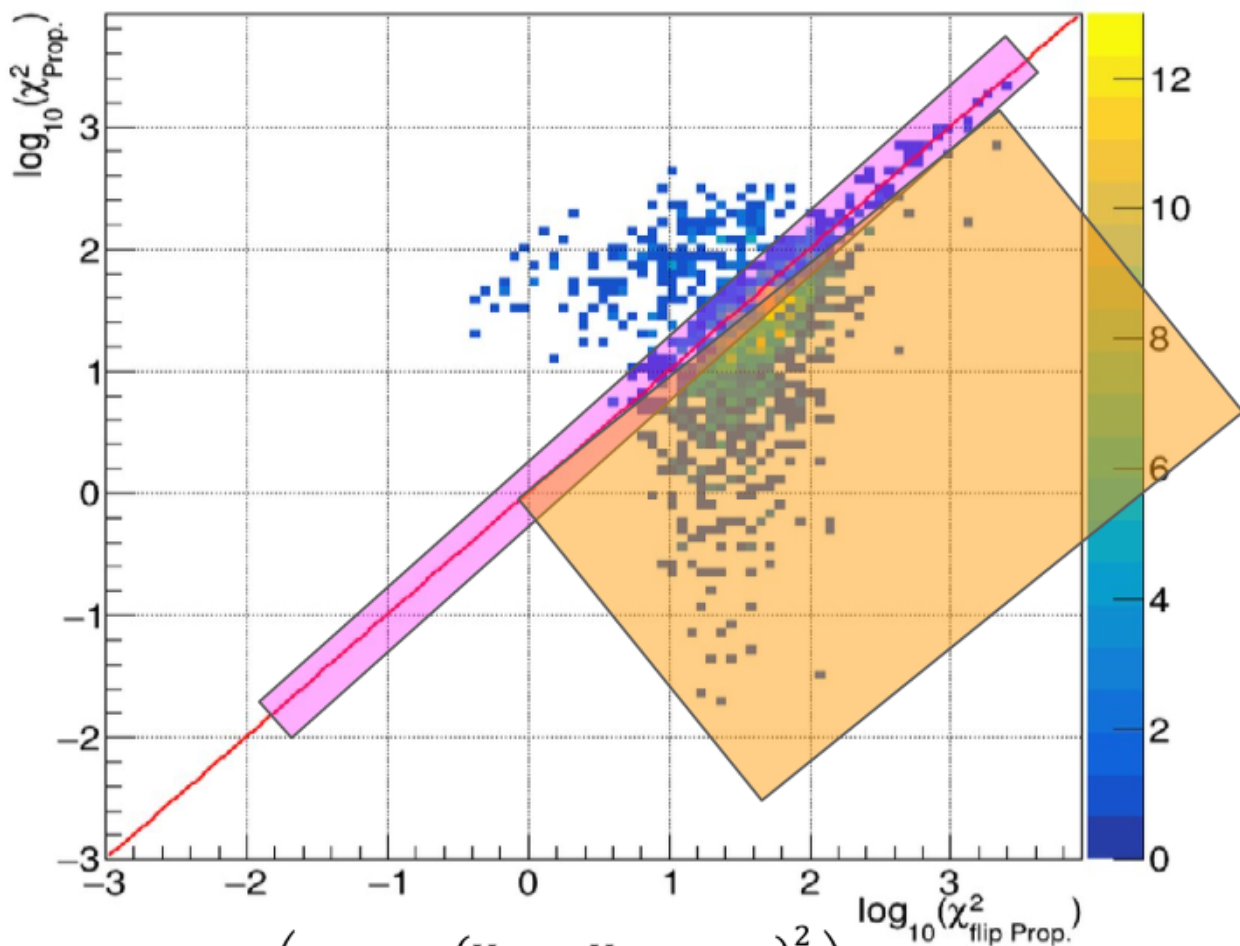
N_FlipProp_bin_(104.337868_130.184189)_xy



$$\sigma_{\text{hit}}^2 = 15 \mu\text{m}$$

$R_{inner} < 0$ (no secondaries within inner tracker)

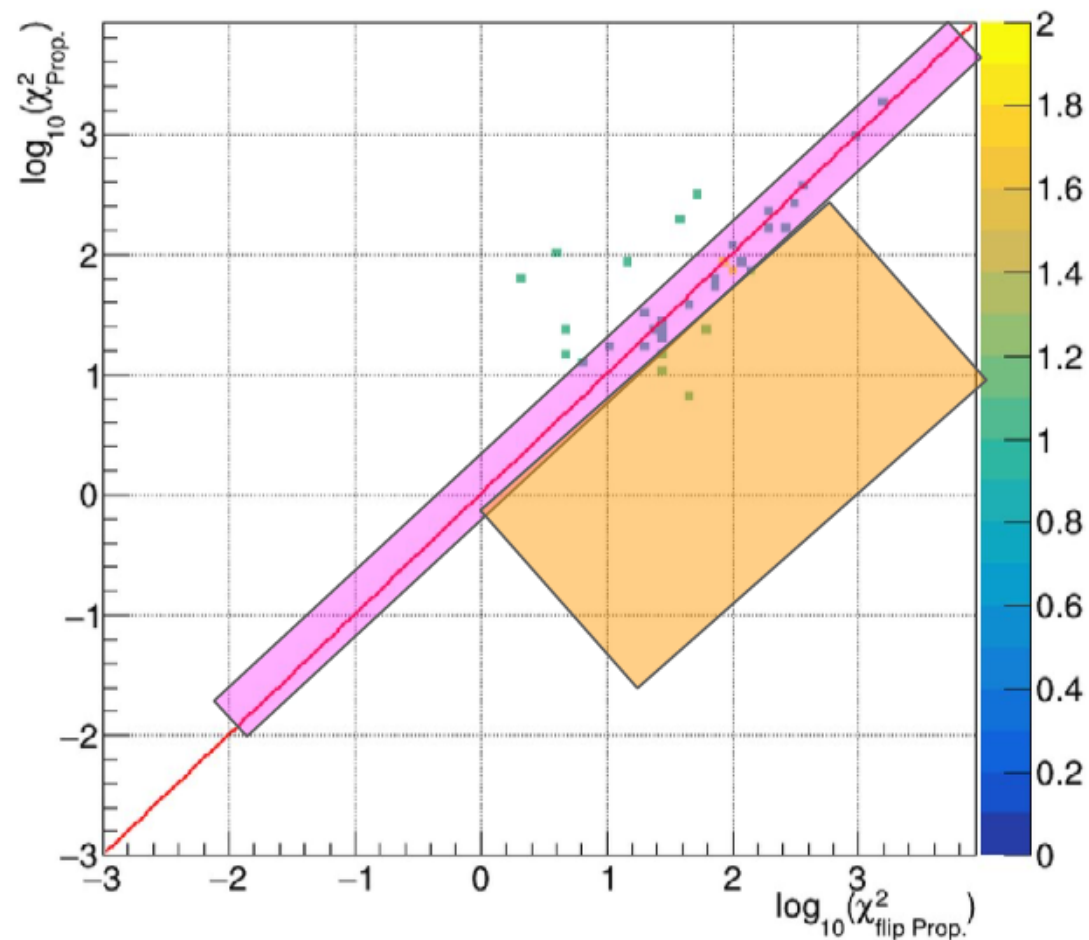
N_FlipProp_bin_(104.337868_130.184189)_xy



$$\chi_{prop}^2 = \frac{\left(\sum_{i-layer} \frac{(Y_{MC} - Y_{prop.track})^2}{\sigma_{hit}^2} \right)}{Ndof}$$

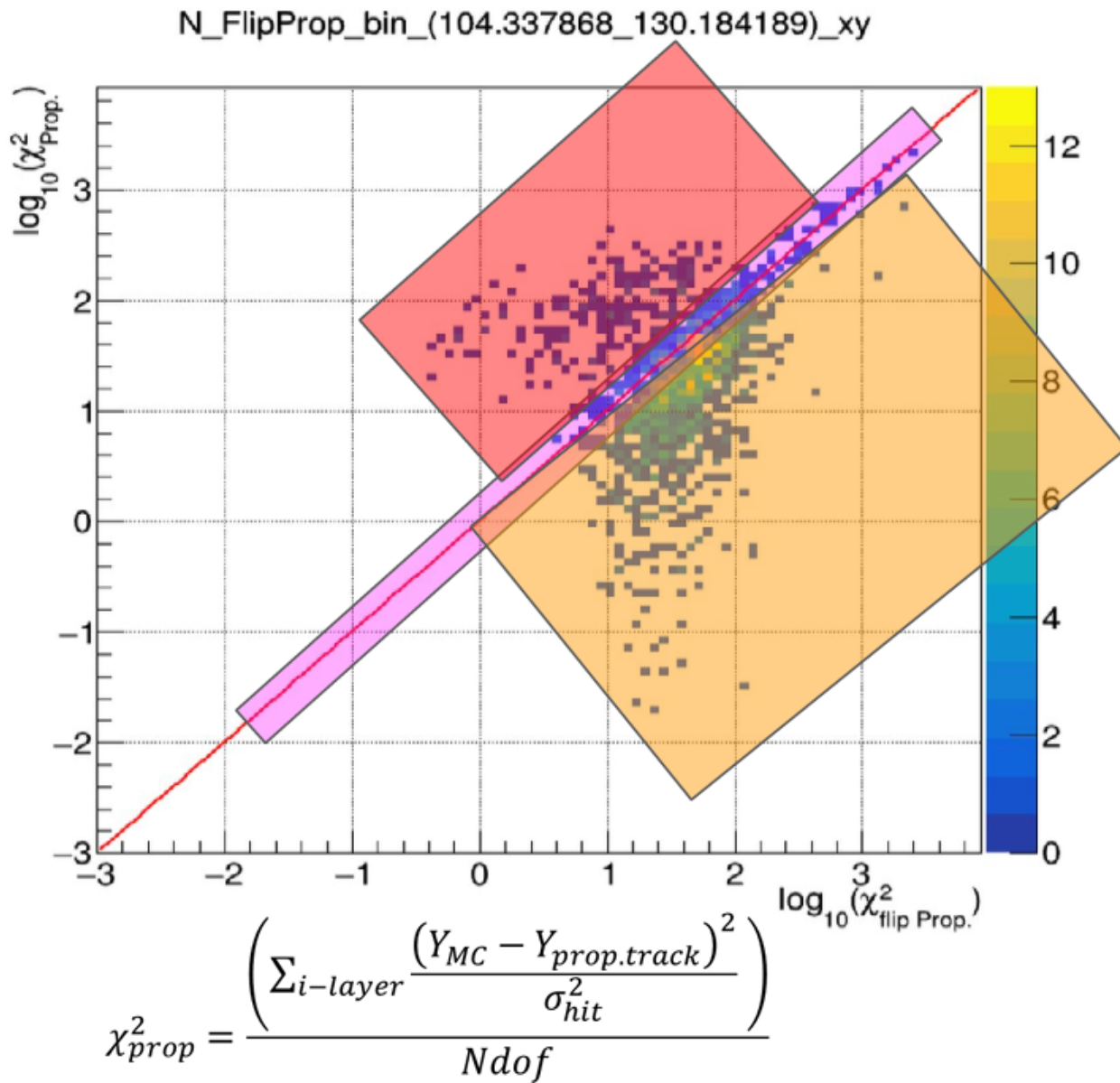
$R_{inner} < 0$ (elastic scattering within inner tracker)

N_FlipProp_bin_(104.337868_130.184189)_xy

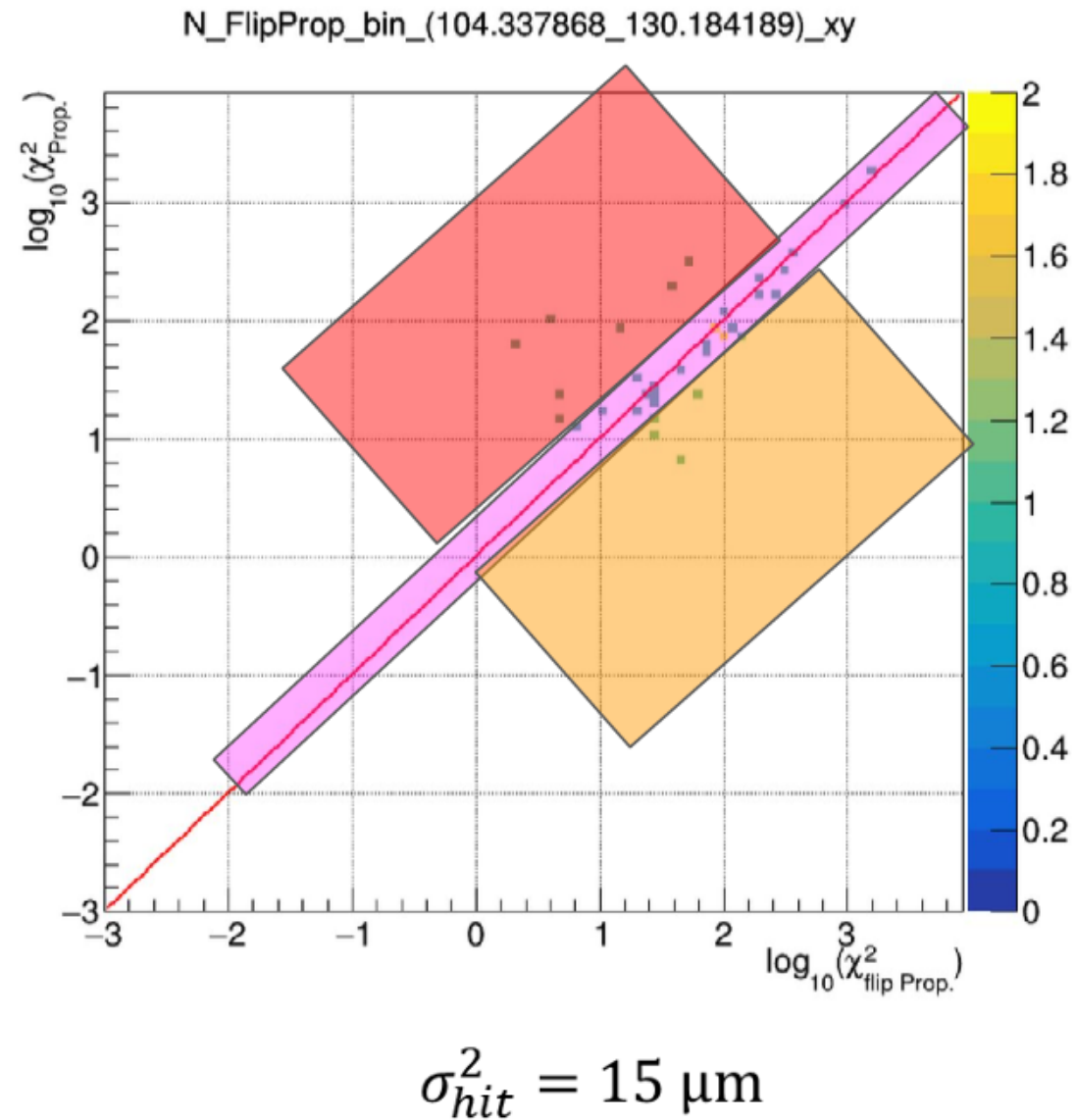


$$\sigma_{hit}^2 = 15 \mu m$$

$R_{\text{inner}} < 0$ (no secondaries within inner tracker)



$R_{\text{inner}} < 0$ (elastic scattering within inner tracker)



Tools to reduce *He* background 1

(supervised learning)

A Fully Connected Neural Network (FCNN) classifier to characterise the *He* background

- Use the classes previously defined as labels for supervised training
- The Monte Carlo has been weighted using published 7.5 *He* flux.
- Sample composition (**14 % Spillover, 2.5% El. Scat., 53% Had. Inel., 30% Other**)
- **Training** sample ($1.78 \cdot 10^5$) and **validation** sample ($0.76 \cdot 10^5$) events
- Choose variables with good data-MC agreement as input features:

Time Of Flight:

- TOF on-time clusters ($\times 4$)

Inner Tracker:

- TRK min feet-distance ($\times 4$)
- TRK Y cluster ($\times 7$)
- TRK Max cluster distance ($\times 7$)
- TRK track hit |Y| ($\times 7$)
- TRK NormEdep2Y* ($\times 7$)

Anti-Coincidence system:

- ACC counters

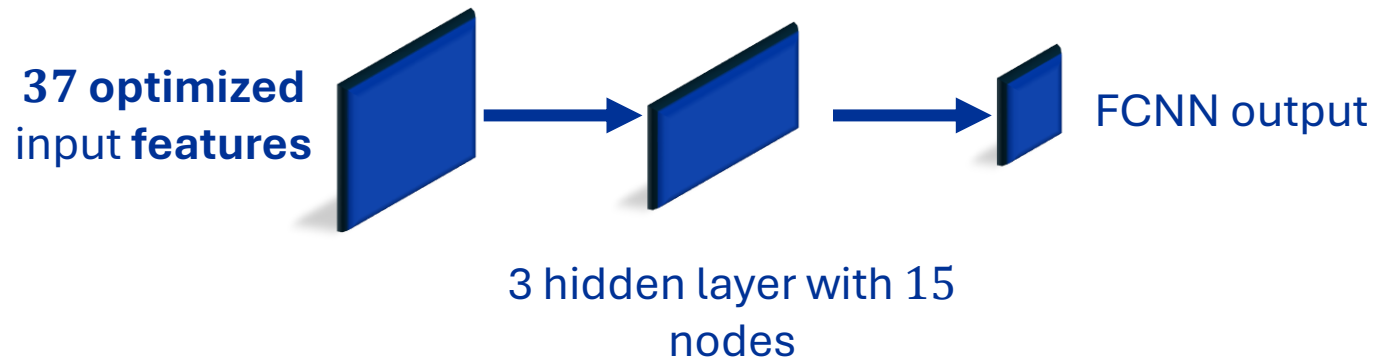
$$*\text{NormEdep2Y} = \frac{\text{Track } E_{dep} Y}{\text{Cluster } E_{dep} Y (2 \text{ cm from Track hit}) + \text{Track } E_{dep} Y}$$

Total number of input features = 37

Fully Connected Neural Network (FCNN) structure

- **FCNN structure:**

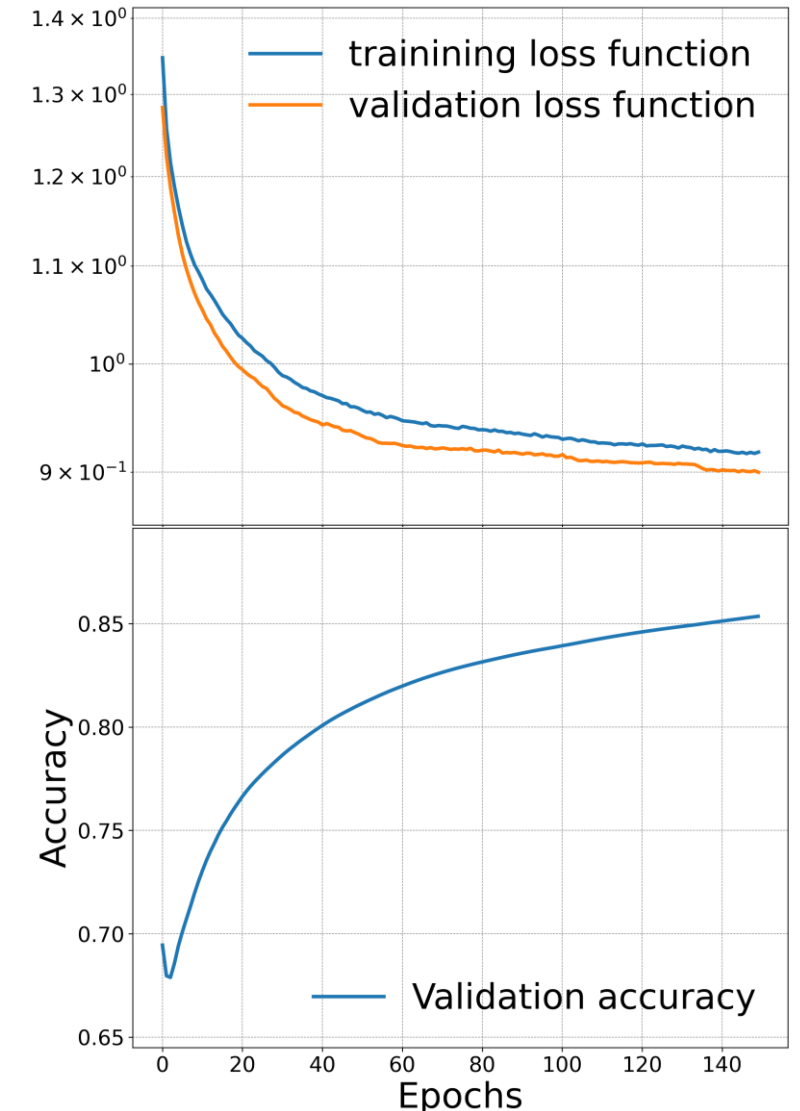
- PyTorch
- Four linear layers: [37, 15, 10, 4]
- Activation functions: ReLu, Softmax (last layer)



- **FCNN hyperparameters:**

- Optimizer: Adam
- Learning rate: $5.0 \cdot 10^{-4}$
- Batch size: $7.0 \cdot 10^2$
- Drop-out: $1.0 \cdot 10^{-1}$
- Loss function: Cross Entropy

Training sample ($1.78 \cdot 10^5$) and validation sample ($0.76 \cdot 10^5$) are unbalanced



Discriminants

- The network returns a vector of four elements.
- Each element corresponds to the probability that the current event belongs to one of the four classes:

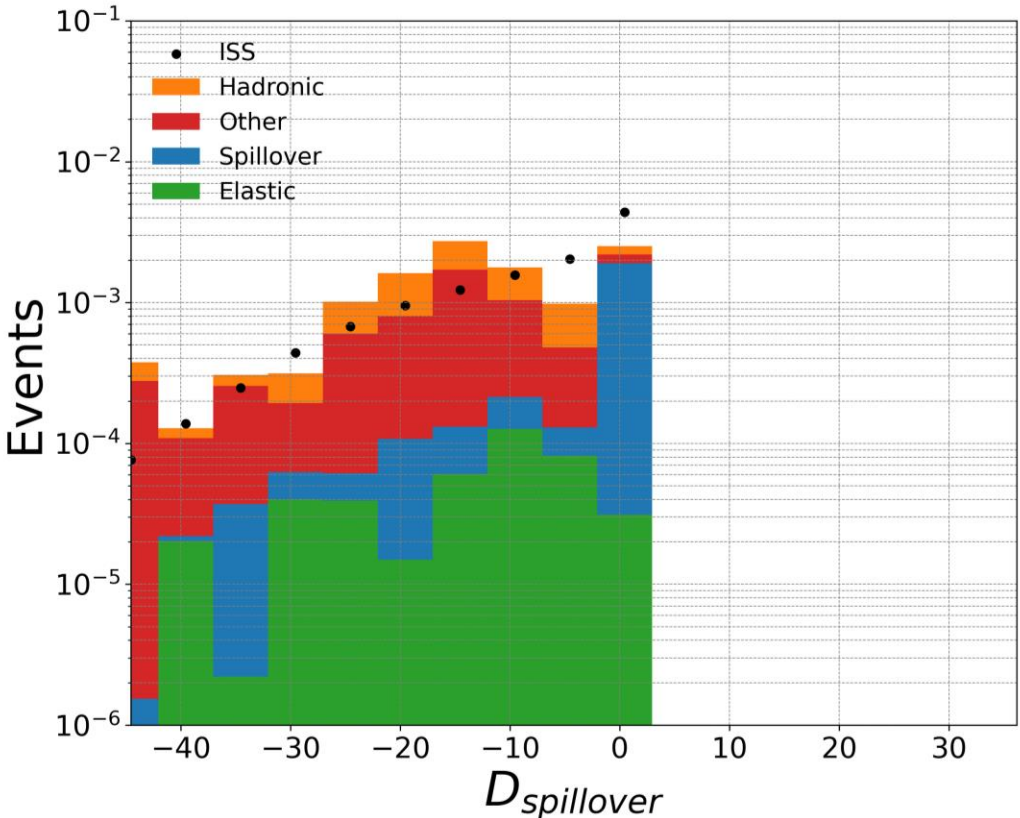
$$\text{FCNN output} = (p_{\text{spillover}}, p_{\text{Had.inel}}, p_{\text{EL.Scat.}}, p_{\text{Other}})$$

- The fraction of each class is defined as: $f_{\text{Had.inel.}} = \frac{\# \text{Had.inel.}}{\# \text{Spillover} + \# \text{Had.inel.} + \# \text{El.sc.} + \# \text{Other}}$
- The discriminant is defined as

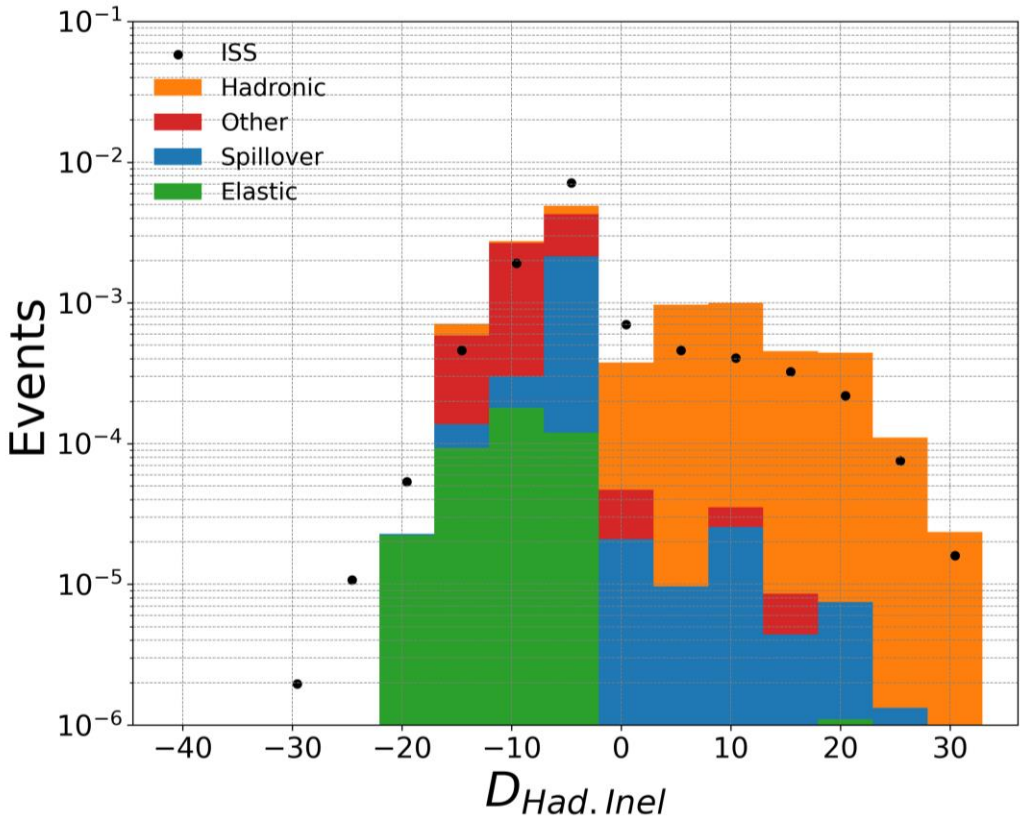
$$D_{\text{Had.inel}} = \log_{10} \left(\frac{p_{\text{Had.inel}}}{f_{\text{Spillover}} \cdot p_{\text{Spillover}} + f_{\text{EL.sc.}} \cdot p_{\text{EL.sc.}} + f_{\text{Other}} \cdot p_{\text{Other}}} \right)$$

- **Applying the Monte Carlo selection (+ RTI cuts and NO cutoff) to data: $1.25 \cdot 10^5$ events with $R_{\text{inner}} < 0$**

Discriminants for spillover and Had. inel.

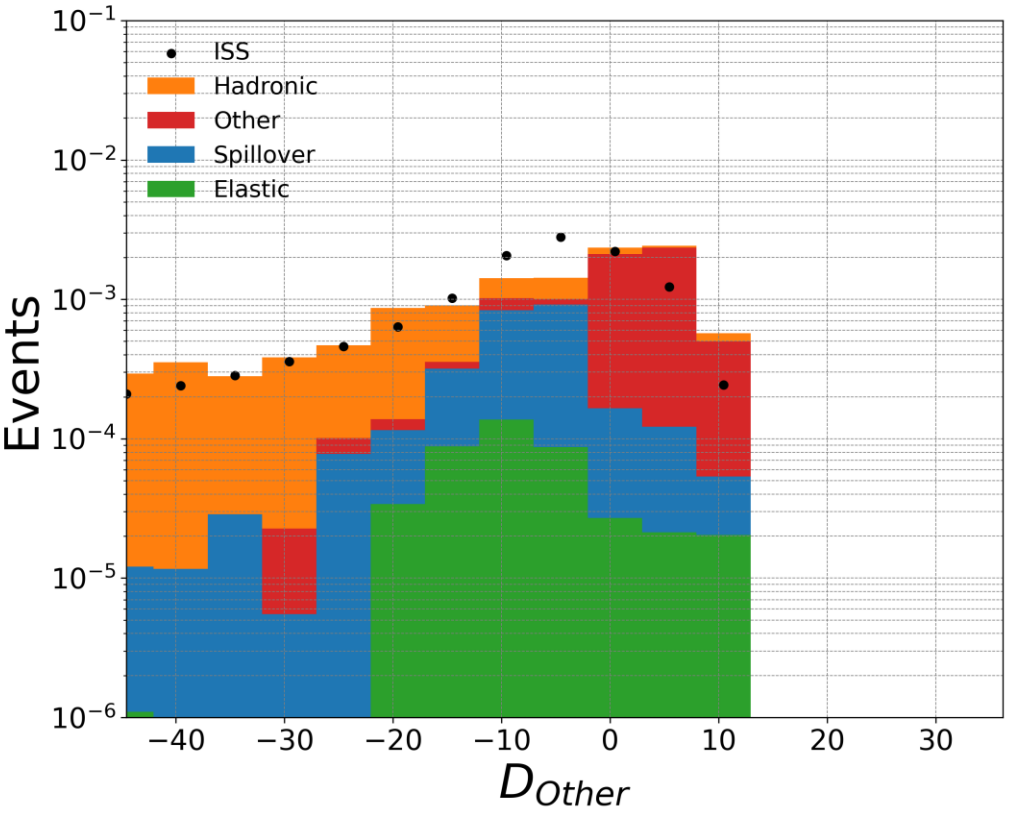
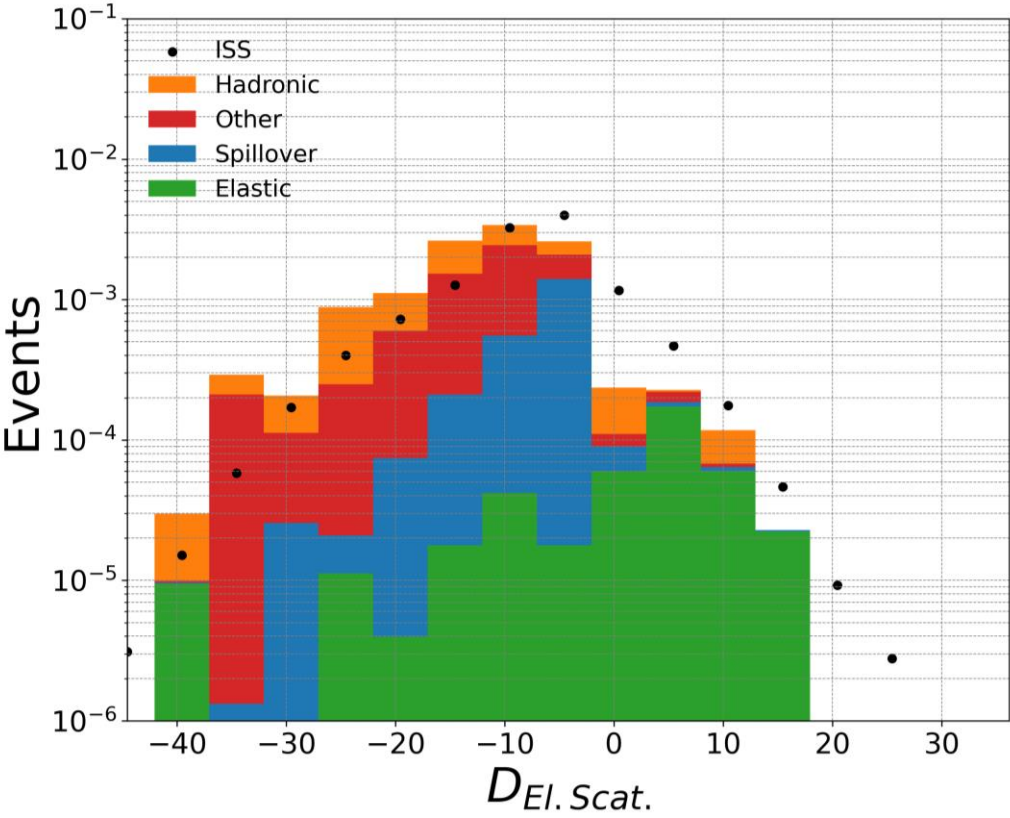


$$D_{Spill} = \log_{10} \left(\frac{p_{spill}}{f_{Inel} \cdot p_{Inel.} + f_{EL.} \cdot p_{EL} + f_{Ot} \cdot p_{Ot}} \right)$$



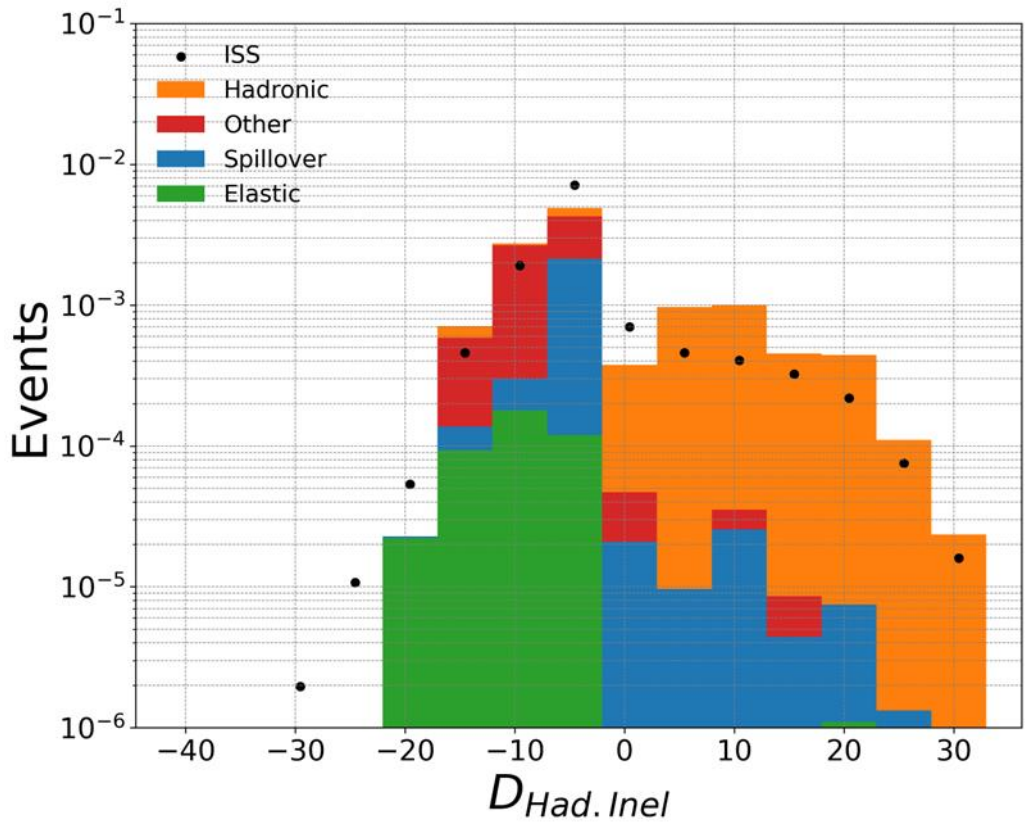
$$D_{EL.} = \log_{10} \left(\frac{p_{Inel.}}{f_{spill} \cdot p_{Spill} + f_{EL} \cdot p_{EL} + f_{Ot} \cdot p_{Ot}} \right)$$

Discriminant for el. scat. and the fourth class

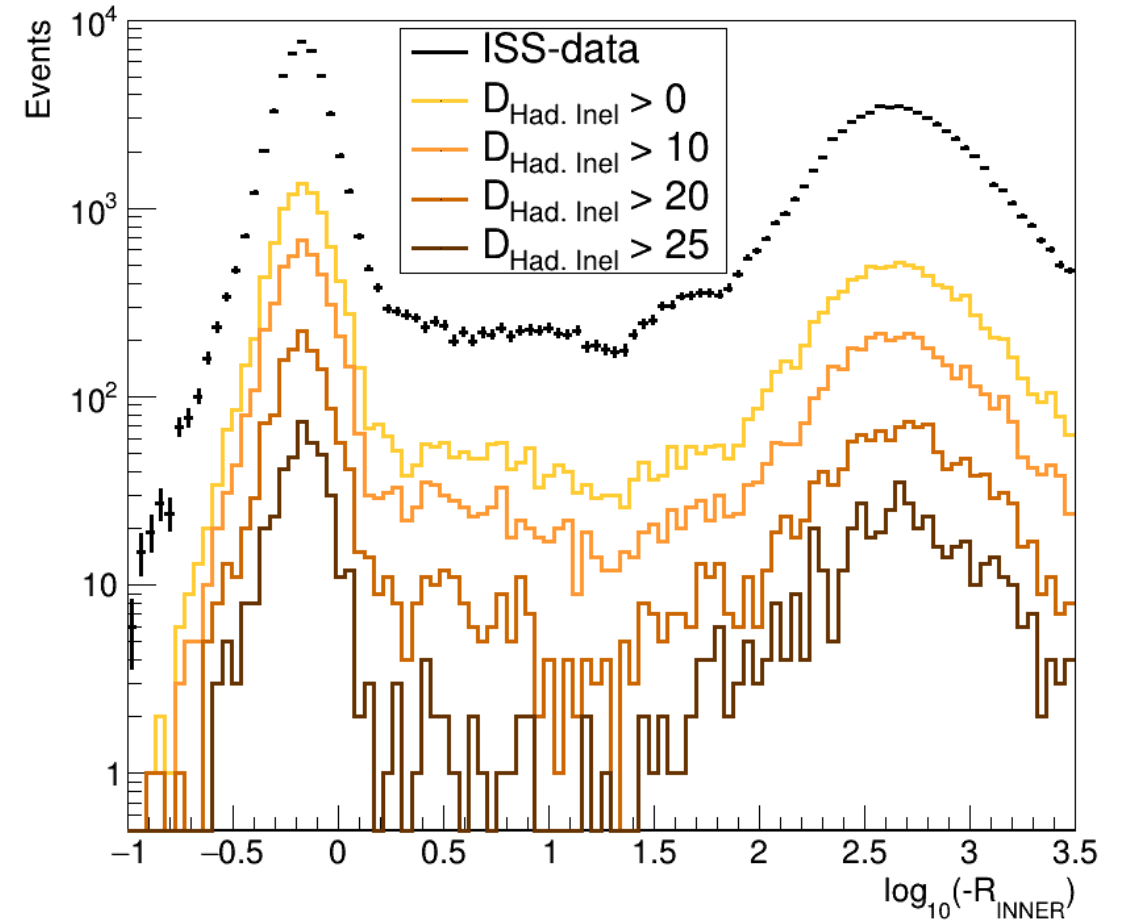


$$D_{El} = \log_{10} \left(\frac{p_{El.}}{f_{Inel} \cdot p_{Inel} + f_{Spill} \cdot p_{Spill} + f_{ot} \cdot p_{ot}} \right)$$

$$D_{ot} = \log_{10} \left(\frac{p_{ot}}{f_{Inel} \cdot p_{Inel} + f_{Spill} \cdot p_{Spill} + f_{El} \cdot p_{El}} \right)$$



$$D_{El.} = \log_{10} \left(\frac{p_{Inel.}}{f_{Spill} \cdot p_{Spill} + f_{El} \cdot p_{El} + f_{Ot} \cdot p_{Ot}} \right)$$



Tools to reduce *He* background 2

(“unsupervised” learning)

Unsupervised learning and autoencoders (AEs)

Why do we need unsupervised learning?

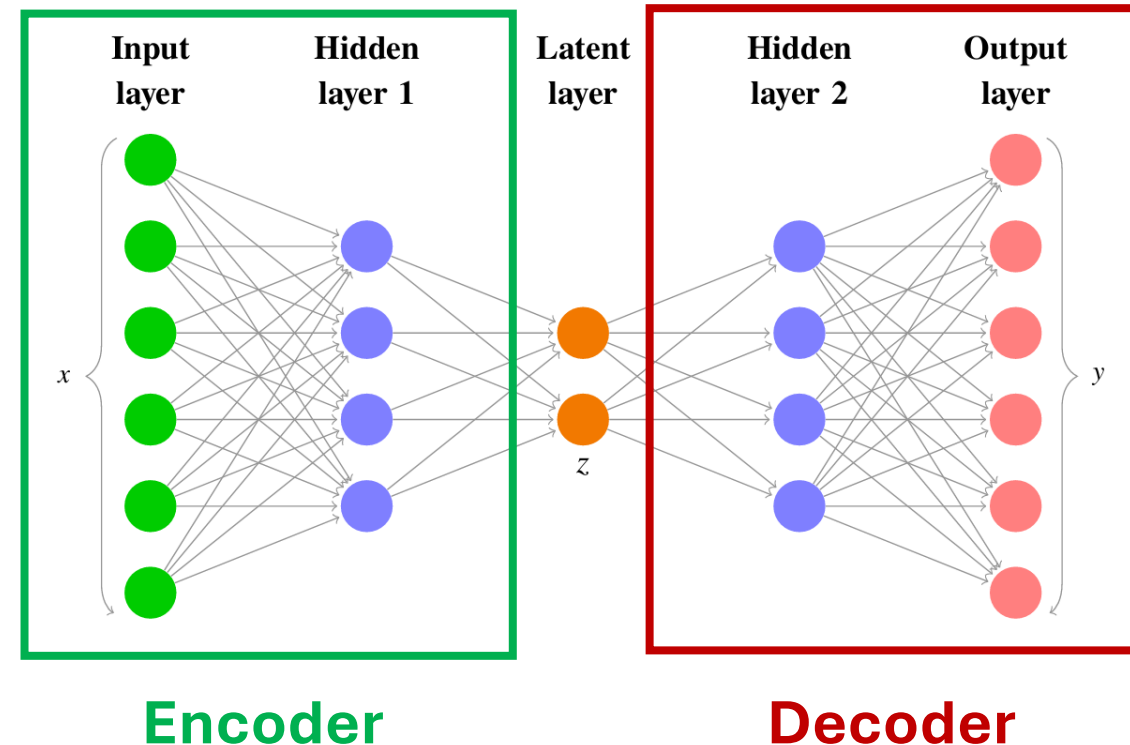
- Unsupervised learning does not need a signal model, it is **model-independent**.

Why autoencoders?

- The network's goal is to reproduce the input, the loss function is the Mean Square Error between the input (\mathbf{X}) and output (\mathbf{Y}) of the AE:

$$L = \frac{(\mathbf{X} - \mathbf{Y})^2}{N_{Features}}$$

- Since signal events are rare, the network should reconstruct them poorly.



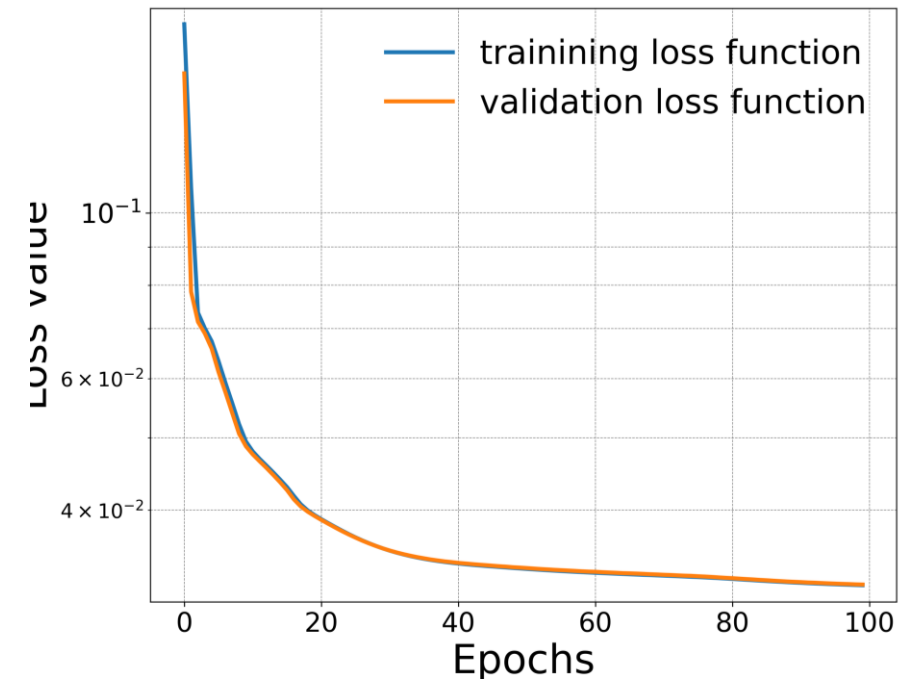
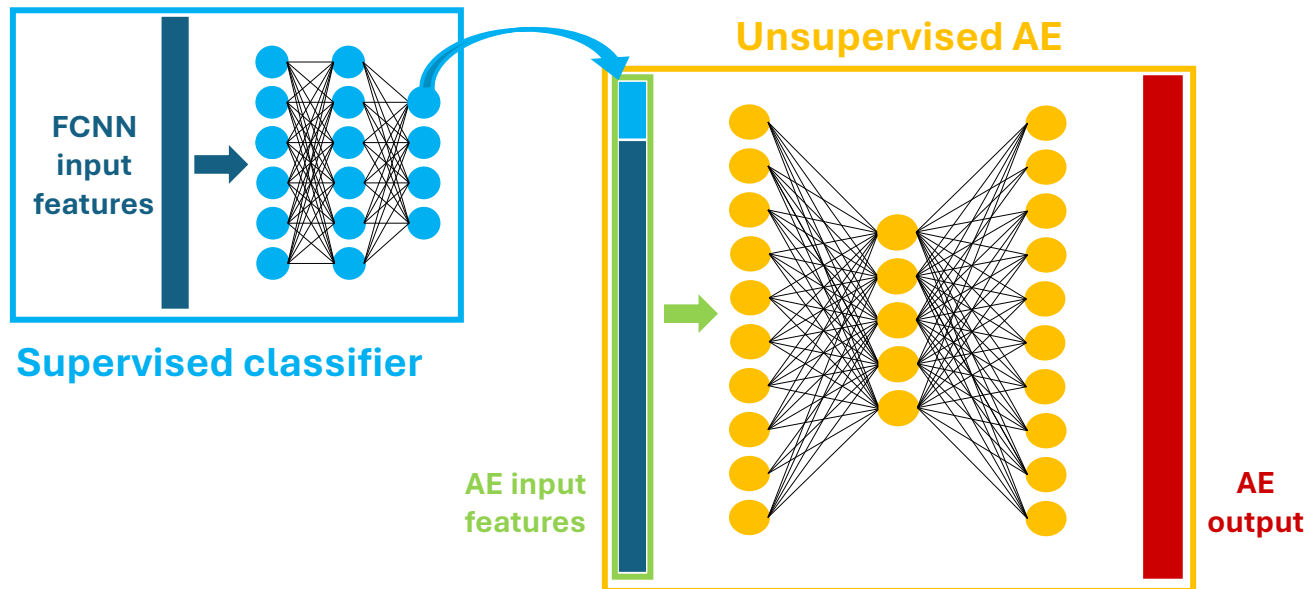
AEs structure and training

Structure:

- [41, 10, 5, 10, 41] (PyTorch)
- Activation functions: ReLu, sigmoid (last layer)
- Optimizer: Adam
- Learning rate: $1.0 \cdot 10^{-4}$
- Batch size: $1.0 \cdot 10^2$
- Loss function: MSE

Training:

- **Training** sample ($0.62 \cdot 10^5$) and **validation** sample ($0.62 \cdot 10^5$) events
- The AE receives as input the 37 features previously described + the 4 FCNN discriminants

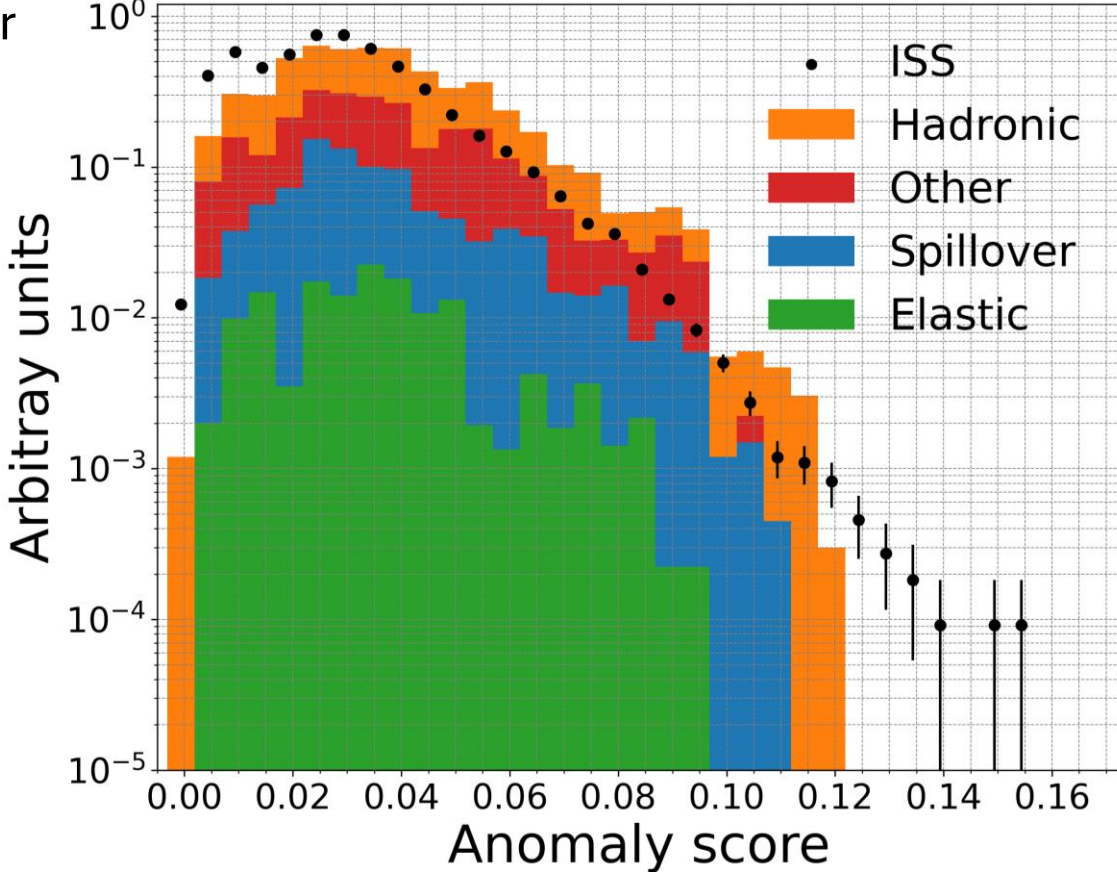
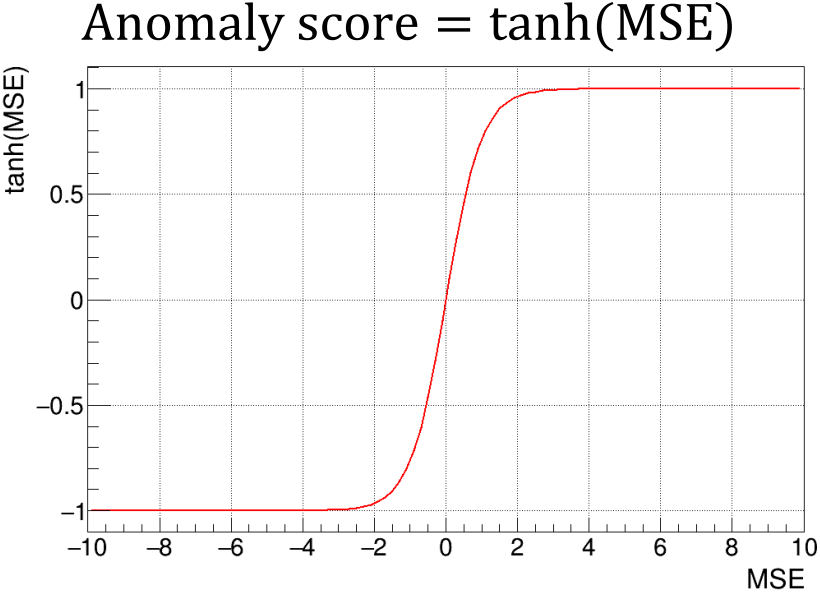


Anomaly detection

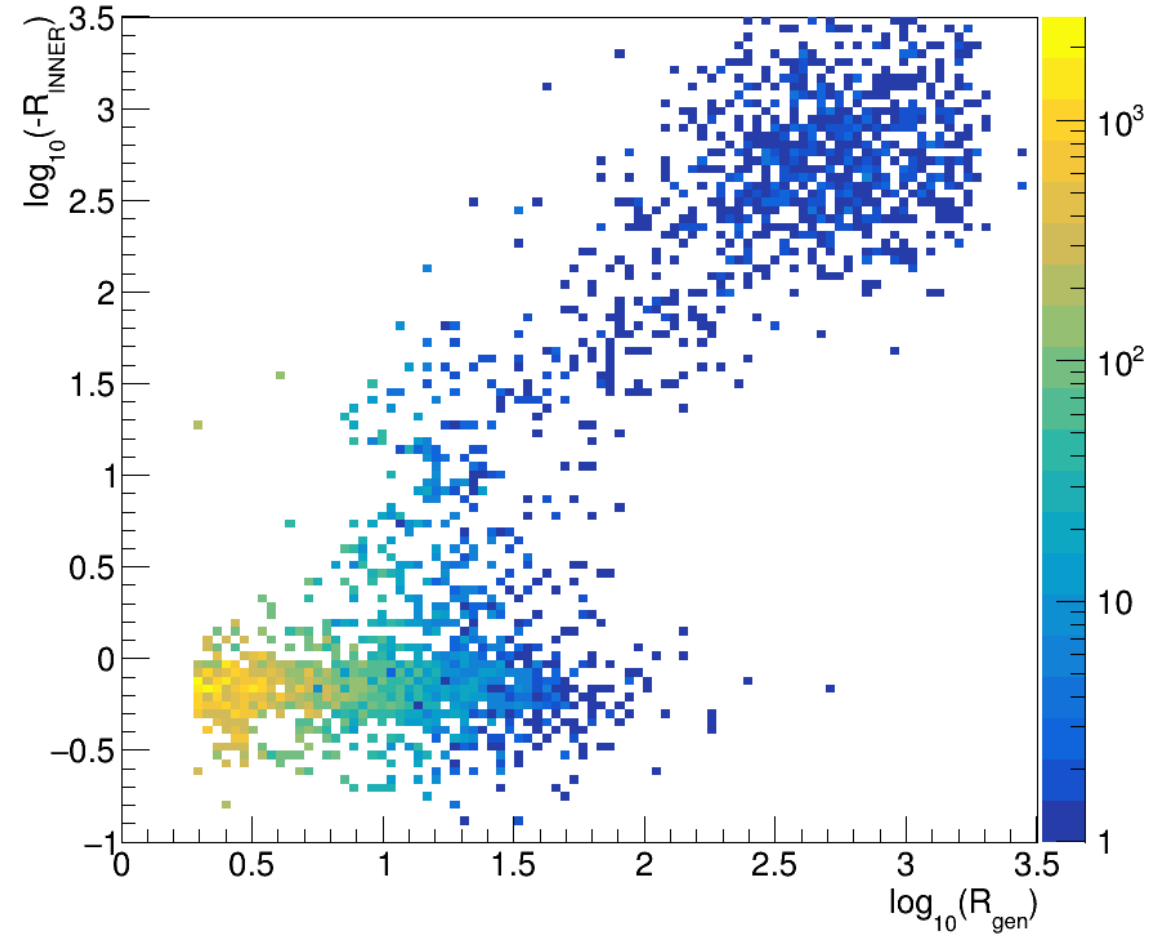
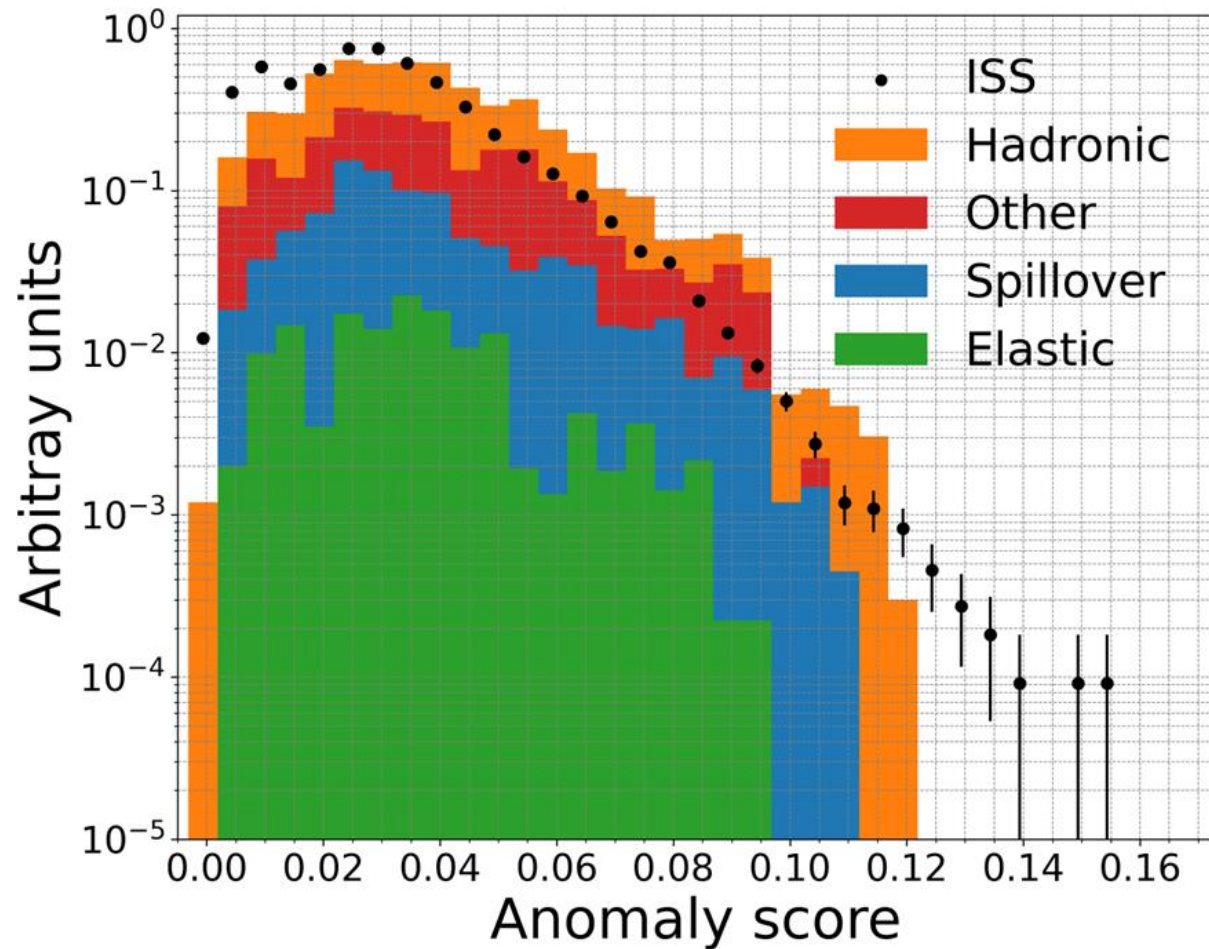
- An anomaly is an event with a high reconstruction error
- The reconstruction error is evaluated using:

$$MSE = \frac{(X - Y)^2}{N_{features}} \quad \begin{matrix} X = \text{input} \\ Y = \text{AE's output} \end{matrix}$$

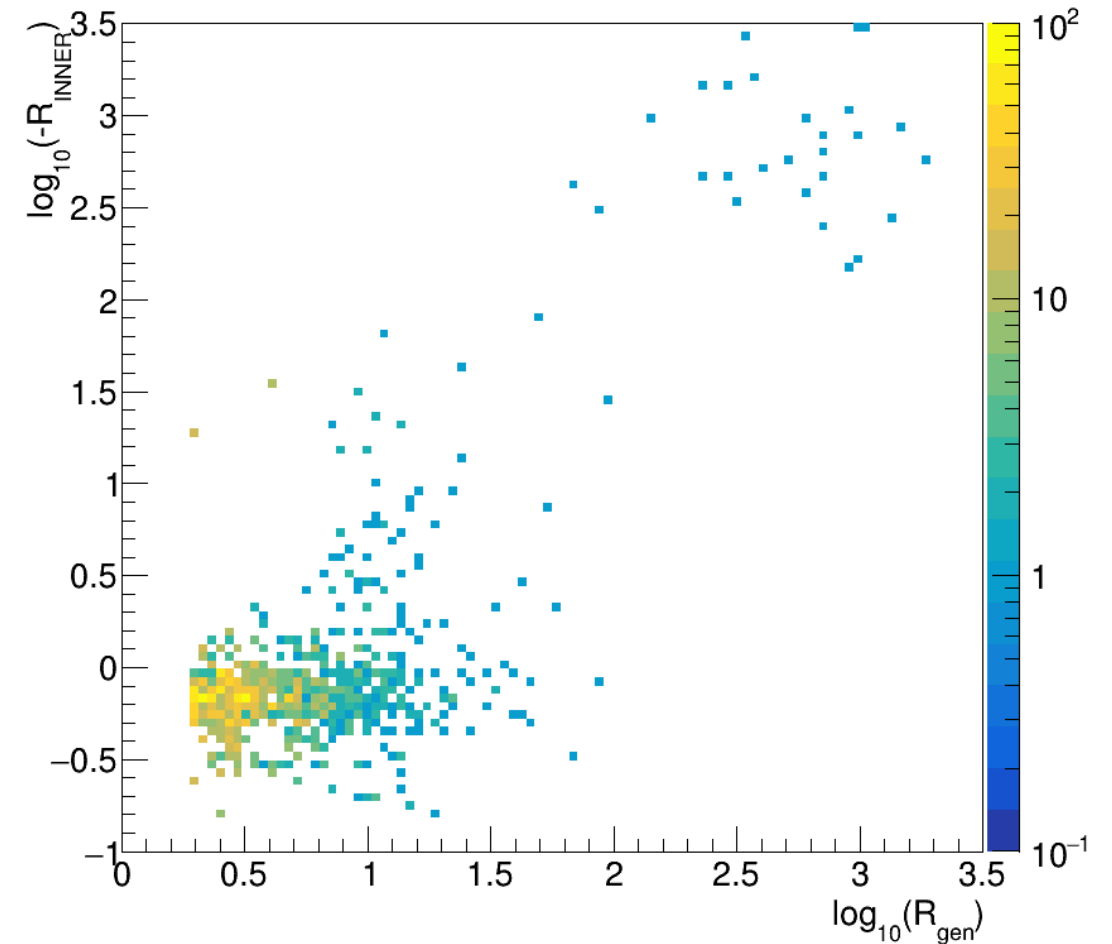
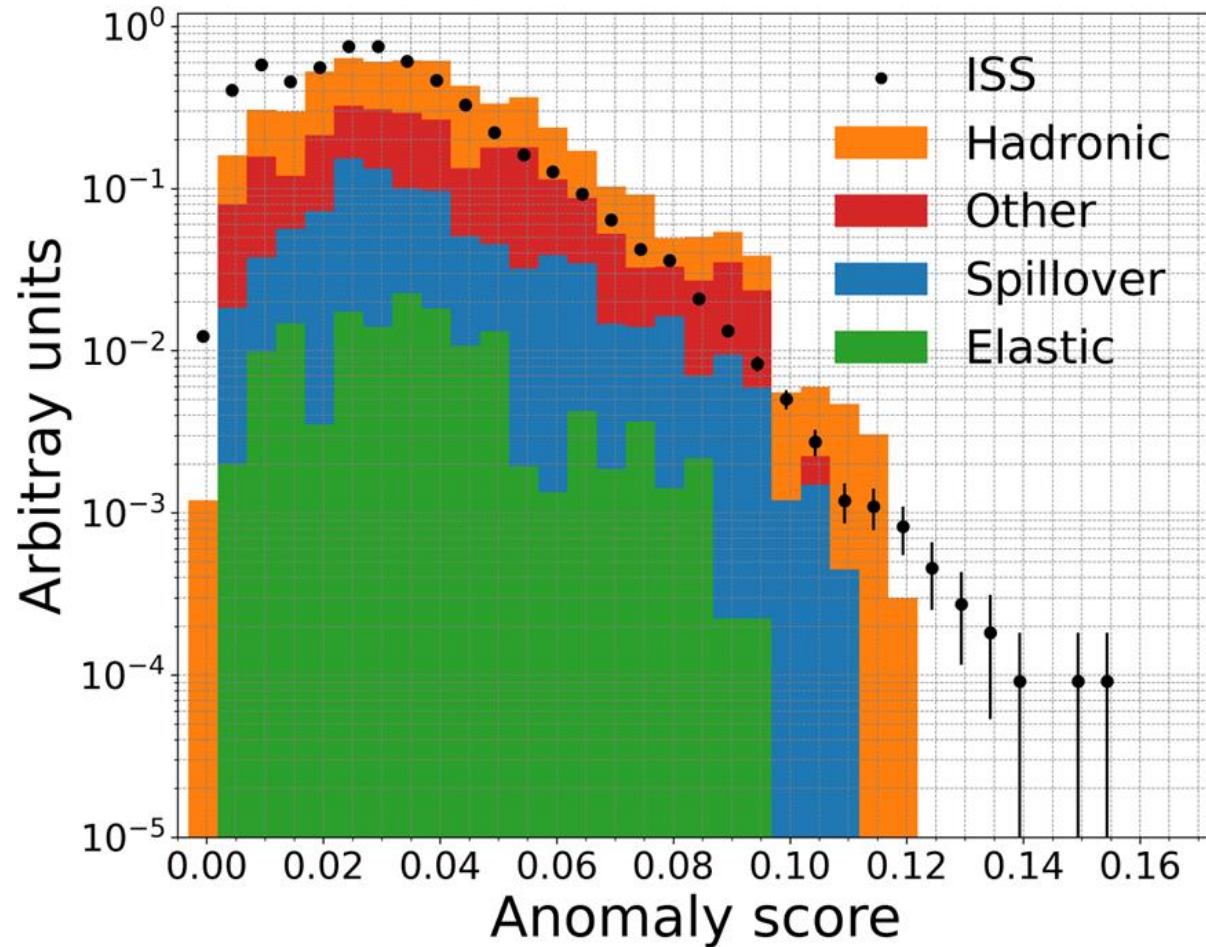
- To have an anomaly score defined between [0,1], the tanh function is used:



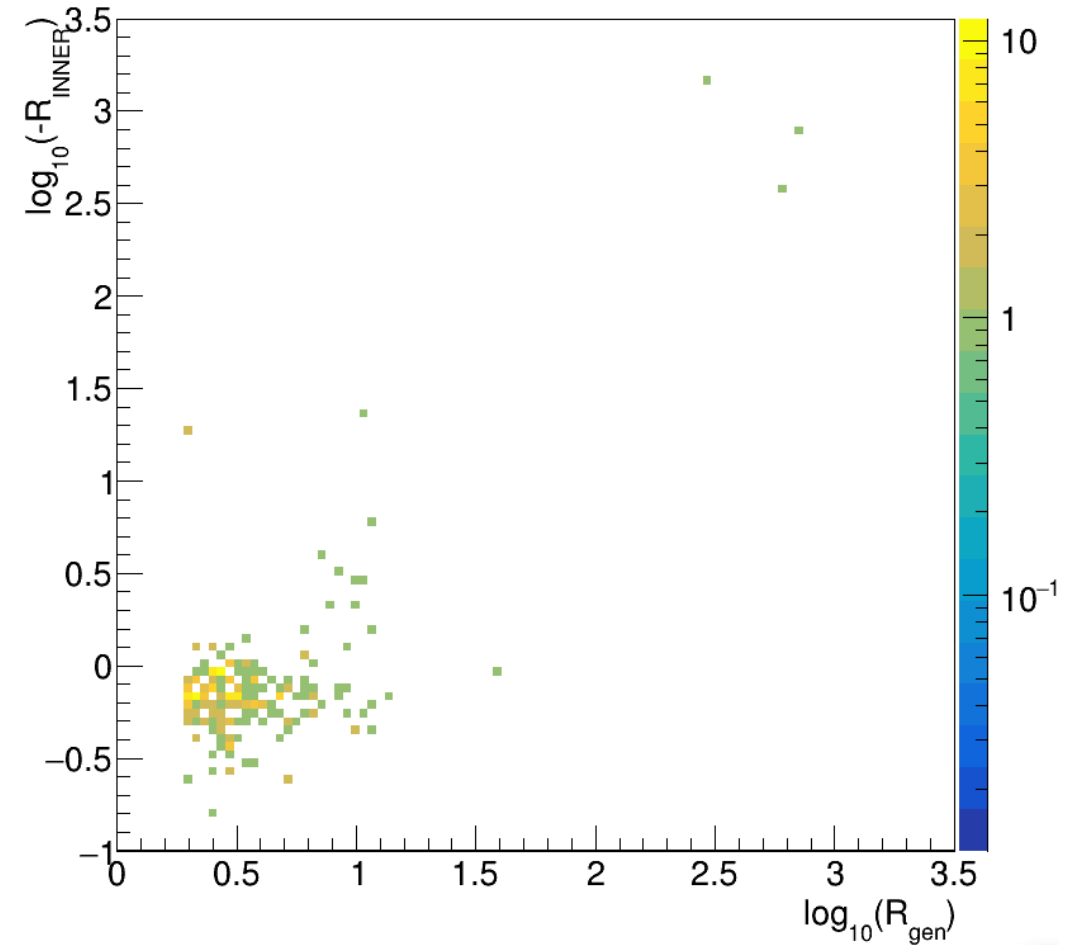
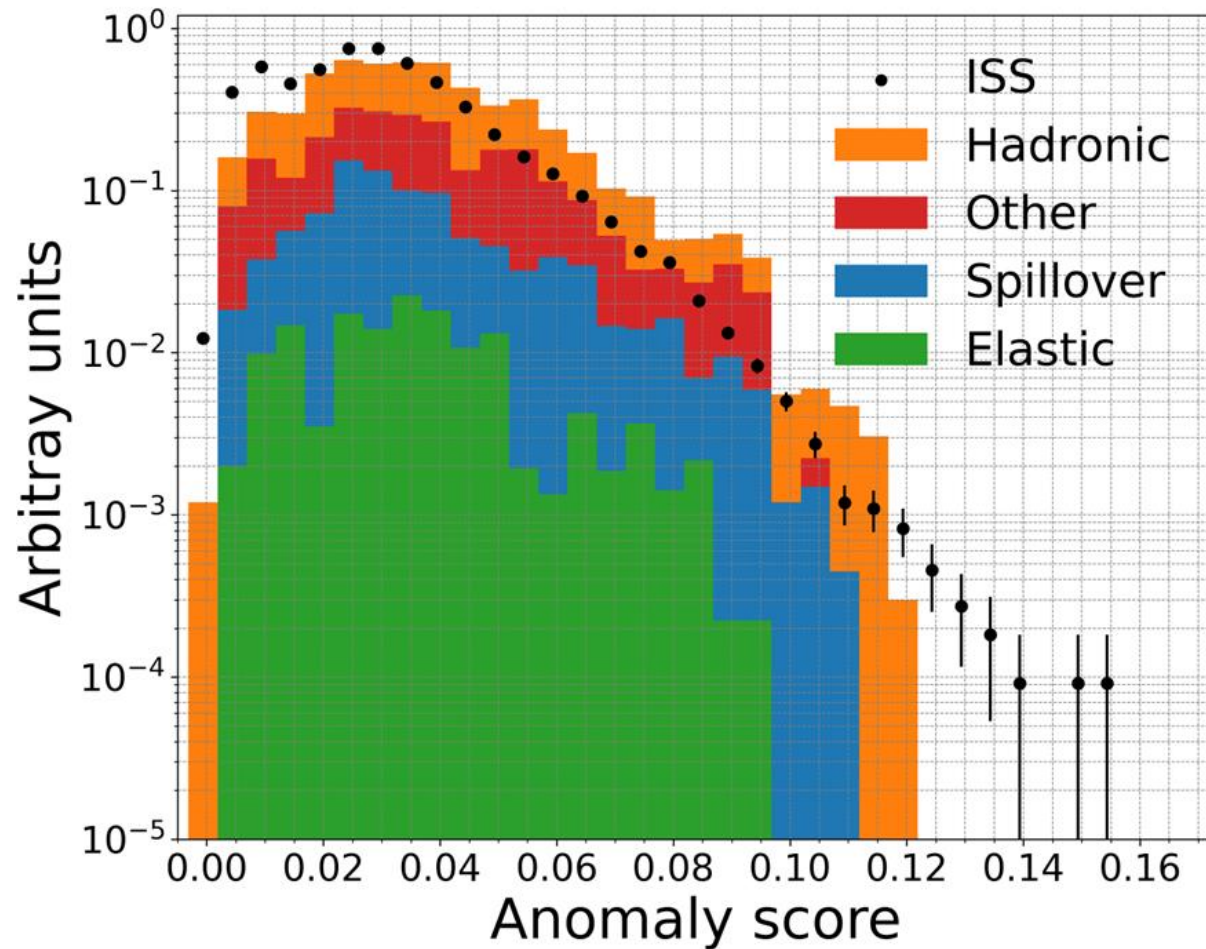
Anomaly score and MC migration matrix

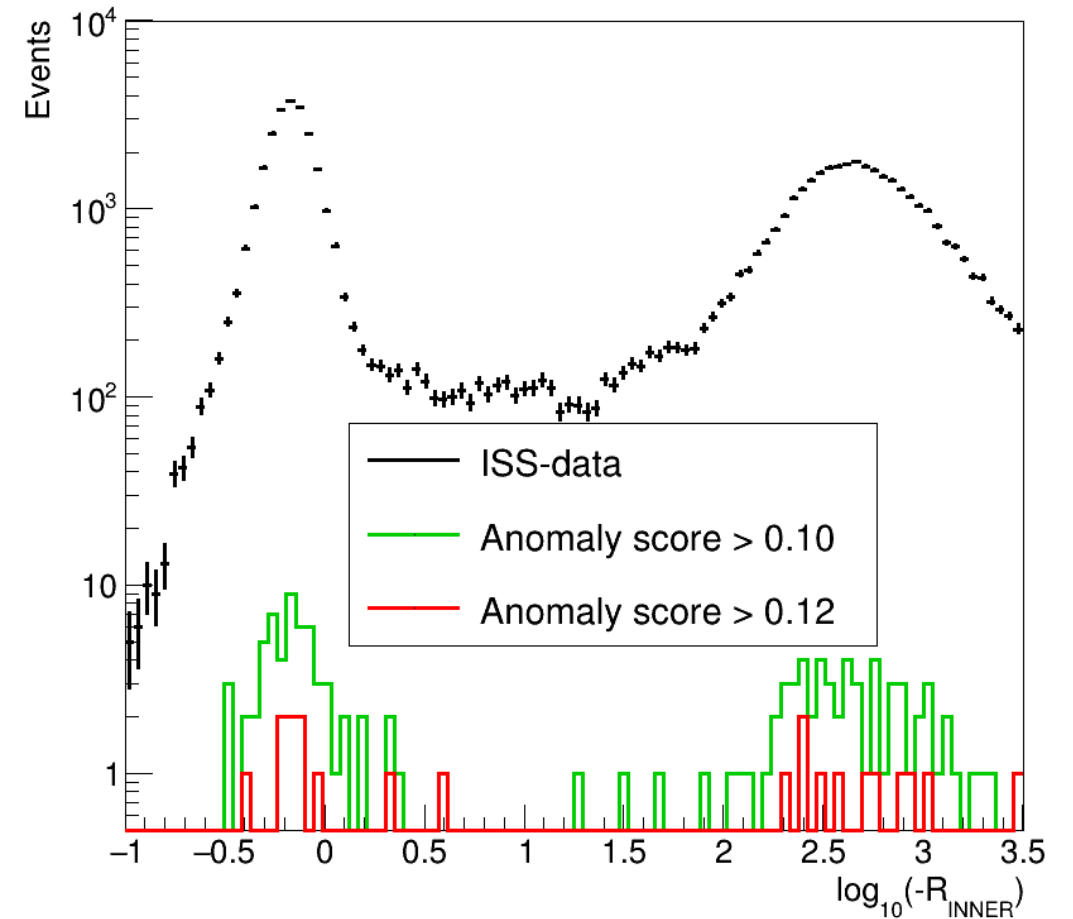
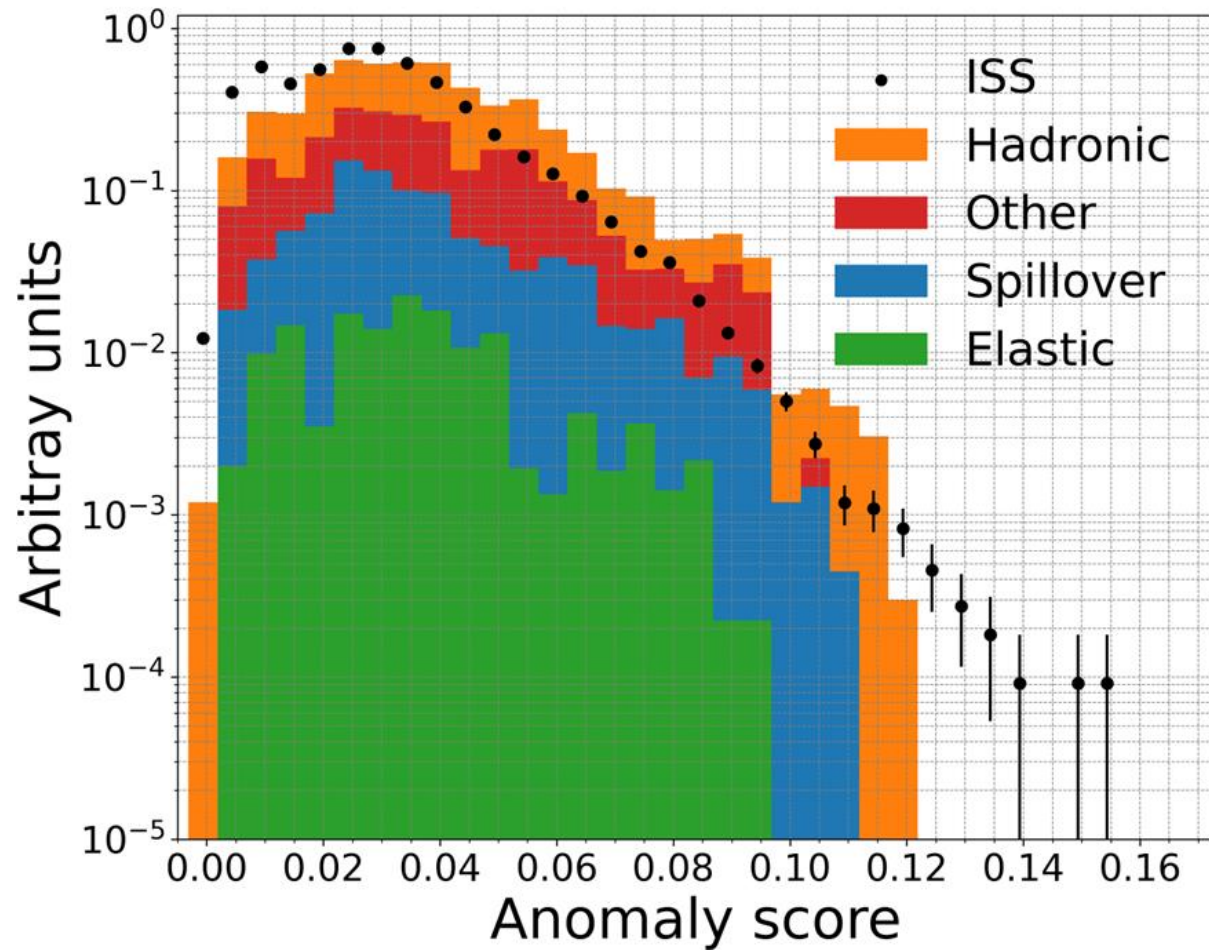


Anomaly score > 0.08 and MC migration matrix



Anomaly score > 0.10 and MC migration matrix





Conclusions

- Labelling charge confusion sources from the Monte Carlo simulation is possible.
- Irreducible background as spillover induces arbitrary choices.
- Unsupervised learning can compensate for the absence of a signal model.
- The combination of a classifier and anomaly detection technique seems promising.

Prospects

- Modify the propagator to take into account energy losses and multiple scattering.
- Investigate new input features to discriminate between spillover and elastic scattering.
- More studies on the rigidity dependence of the input variables.

- Optimize the FCNN classifier and the AE
- Checks on data outliers

Conclusions

- Labelling charge confusion sources from the Monte Carlo simulation is possible.
- Irreducible background as spillover induces arbitrary choices.
- Unsupervised learning can compensate for the absence of a signal model.
- The combination of a classifier and anomaly detection technique seems promising.

Thank you for your attention!

Prospects

- Modify the propagator to take into account energy losses and multiple scattering.
- Investigate new input features to discriminate between spillover and elastic scattering.
- More studies on the rigidity dependence of the input variables.

- Optimize the FCNN classifier and the AE
- Checks on data outliers

Backup

Elastic scattering Monte Carlo

Selection

Selezione comune attualmente usata

IsPhysicsTrigger

$\beta > 0$

TOF hits = 4

Chi2Coo < 4

Track number ≥ 1

charge YJ (inner) $\in [1.7, 2.4]$

inner fiducial volume

charge YJ (L1) $\in [1.6, 3.0]$

track pattern 5/8 (L1-inner)

$\chi_Y^2 < 10$

charge (UTOF) $\in [1.5, 3.0]$

charge (LTOF) > 1.5

($R_{UH}, R_{LH} < 0$) oppure ($R_{UH}, R_{LH} > 0$)

NoCut	15376772	1
RTIGood	15376772	1
RTIIsInSAA(0)	15376772	1
RTLIVETimeFraction(0.5)	15376772	1
IsPhysicsTrigger	7526862	0.4892
BetaPos(0.2)	6383353	0.4148
NTOFBetaClusters(4)	5849648	0.3801
BetaChi2Coo(4)	4476765	0.2907
NTrTracks(1)	4476765	0.2907
HasGBLFitInner	4472418	0.2904
ChargeInnerTrackerYJ(1.7,2.4)	3905926	0.2535
CheckFiducialInner	3311273	0.2149
ChargeLayer1(1.6,3)	2997280	0.1946
IsInsideL1Fiducial	2997274	0.1946
CheckTrackPattern(5)	2680073	0.1740
Chi2Y_GBL_InnerOnly(10)	2573763	0.1672
ChargeUpperTof(1.5,3)	2550871	0.1657
ChargeLowerTof(1.5,30)	2529516	0.1642
HasGBLFitUHInner	2434282	0.1581
HasGBLFitLHInner	2433791	0.1580
SignUHandLH	2379294	0.1545

$R_{inner} < 0 \longrightarrow 696 \text{ eventi}$

Distribution of the scattering angle ($R < 0$)

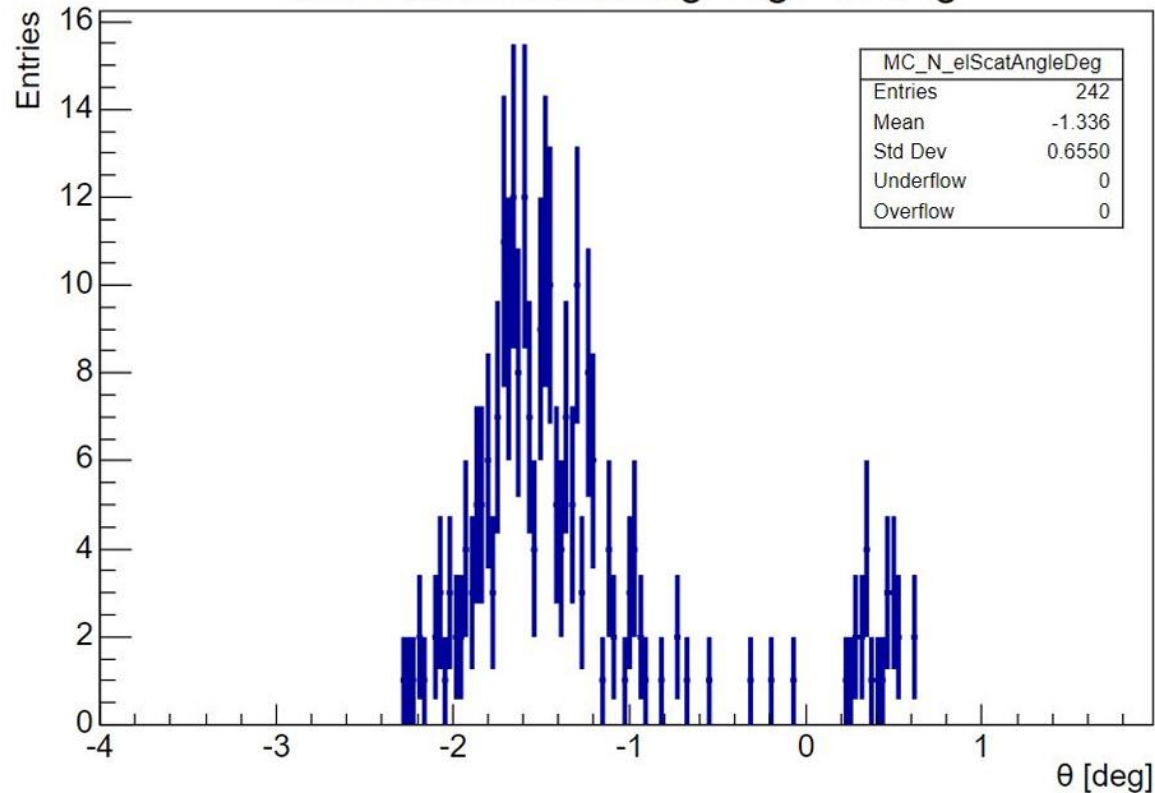
$R_{\text{inner}} < 0$ \longrightarrow 696 eventi

Verità MC per scattering elastic



242 eventi con scattering elastico nell'inner tracker

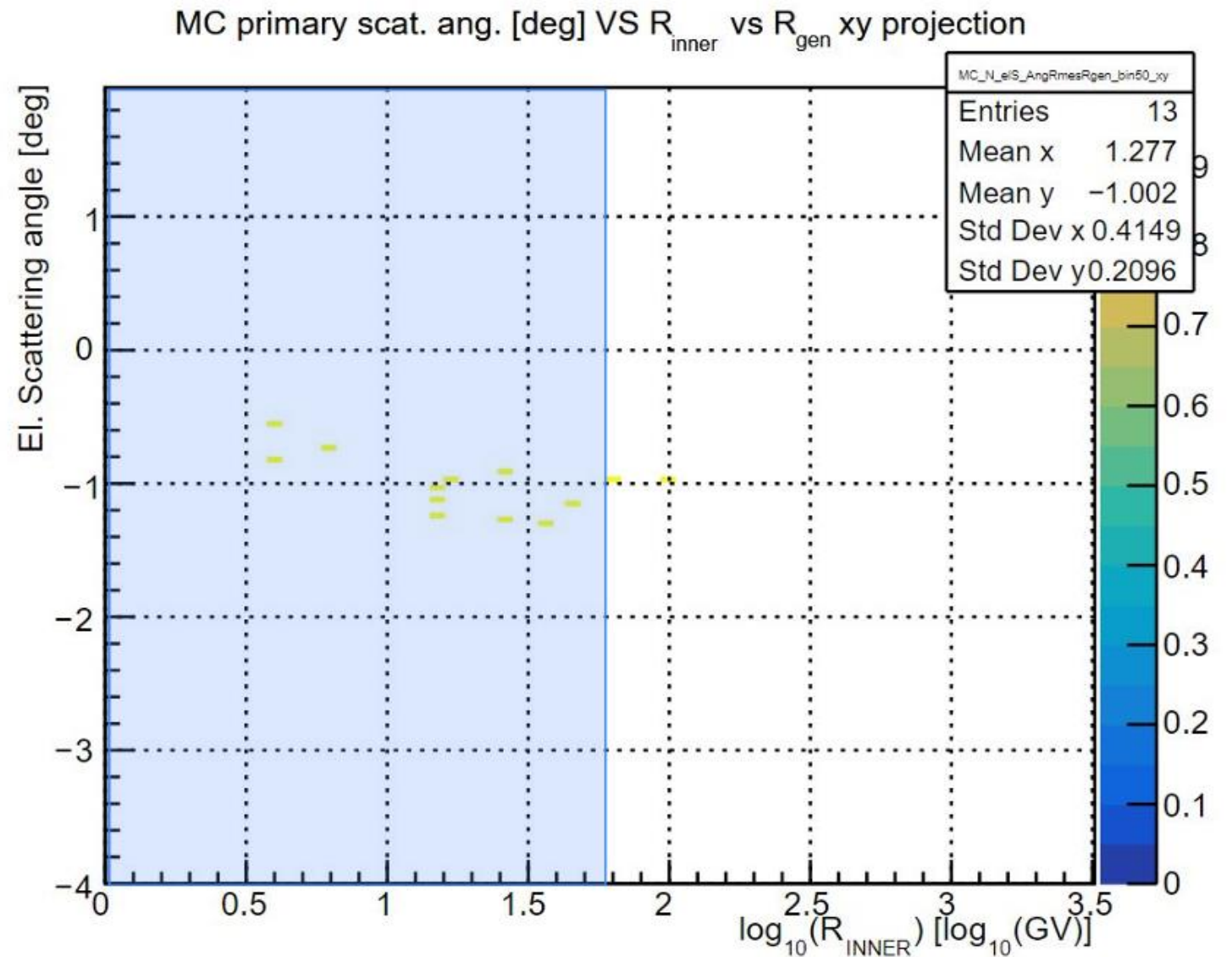
MC elastic scattering angle in deg



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

$$R_{gen} \in [0, 50[$$

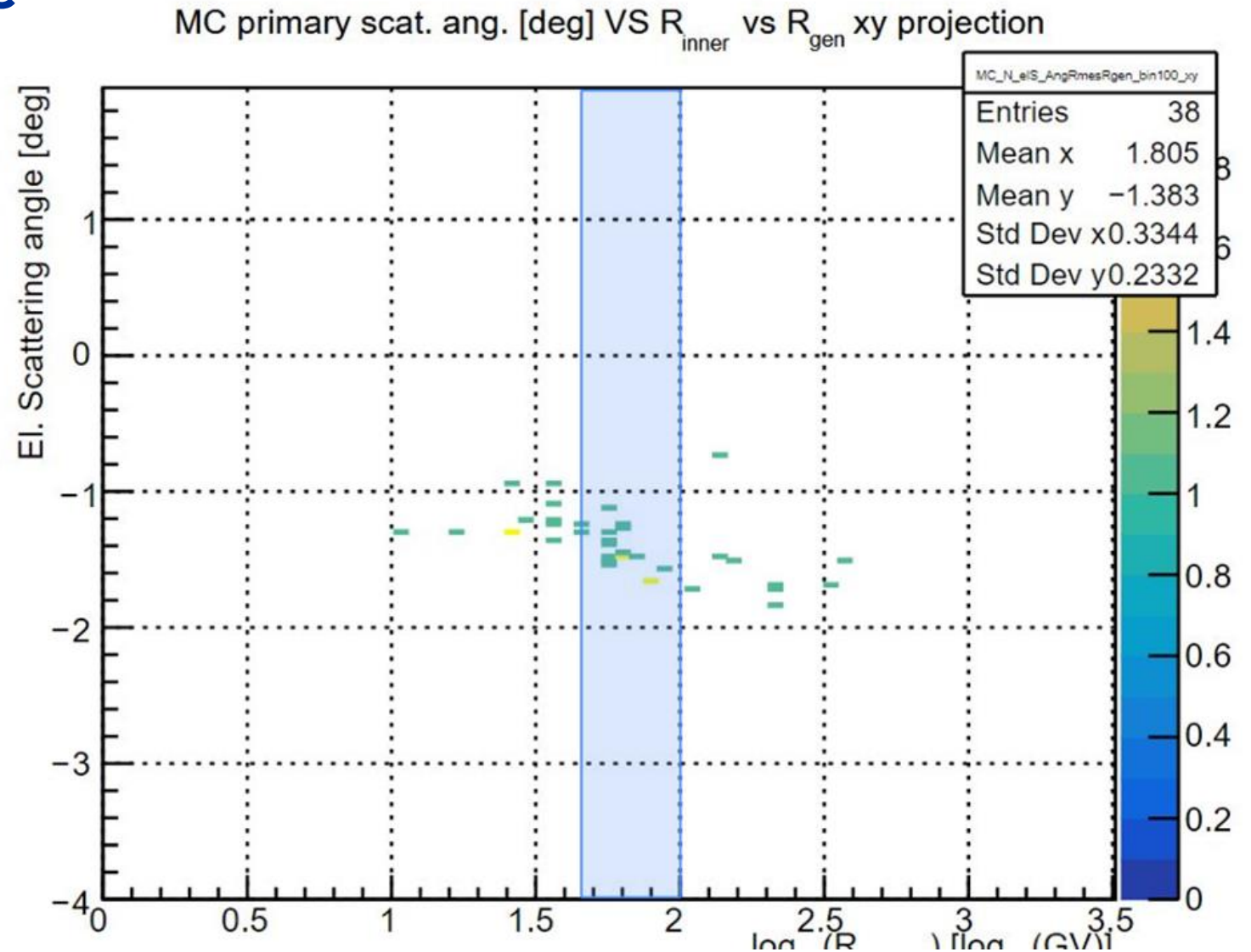
$$\log_{10}(R_{gen}) < 1.67$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

$$R_{gen} \in [50, 100[$$

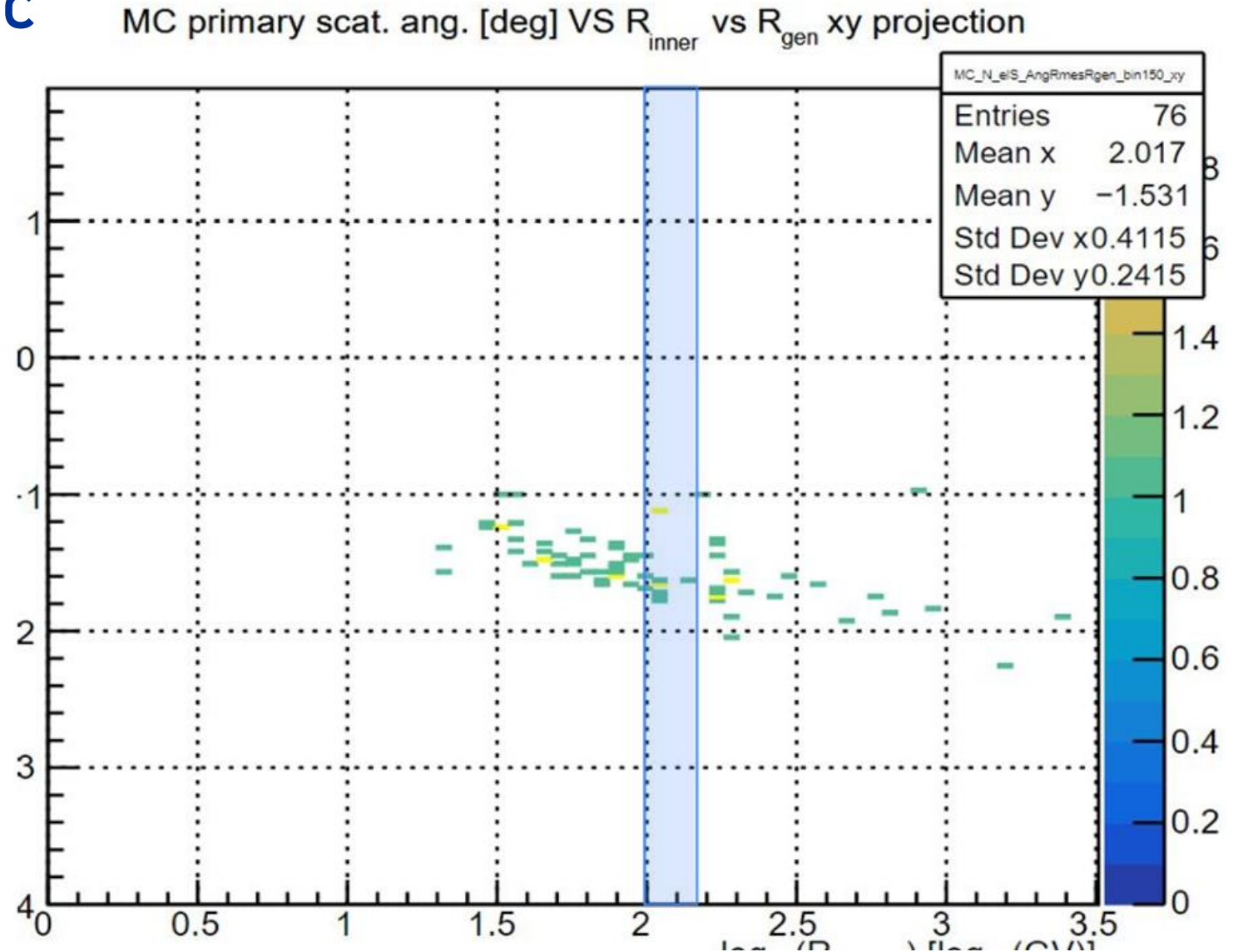
$$\log_{10}(R_{gen}) < 2.0$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

$$R_{gen} \in [100, 150[$$

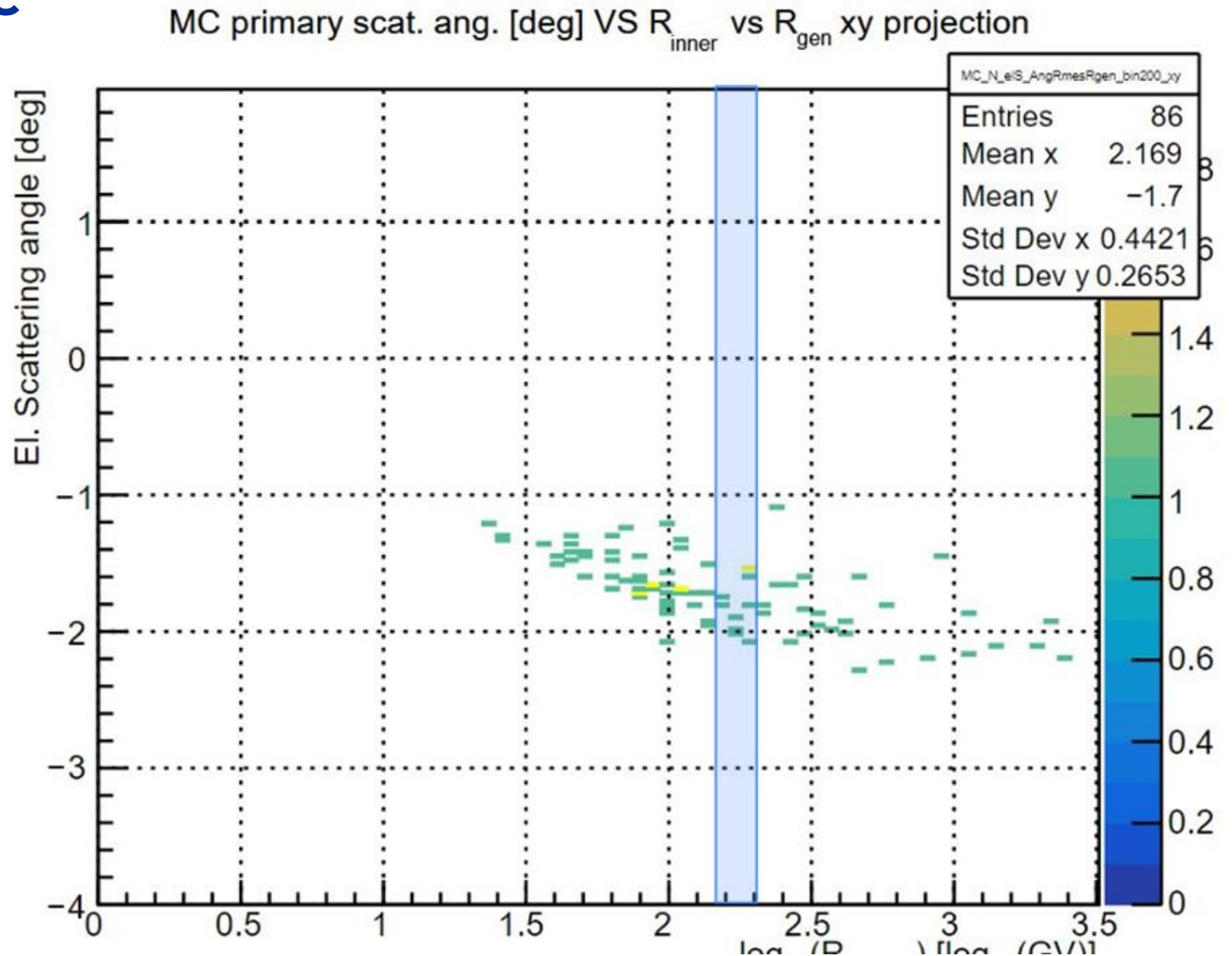
$$\log_{10}(R_{gen}) < 2.18$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

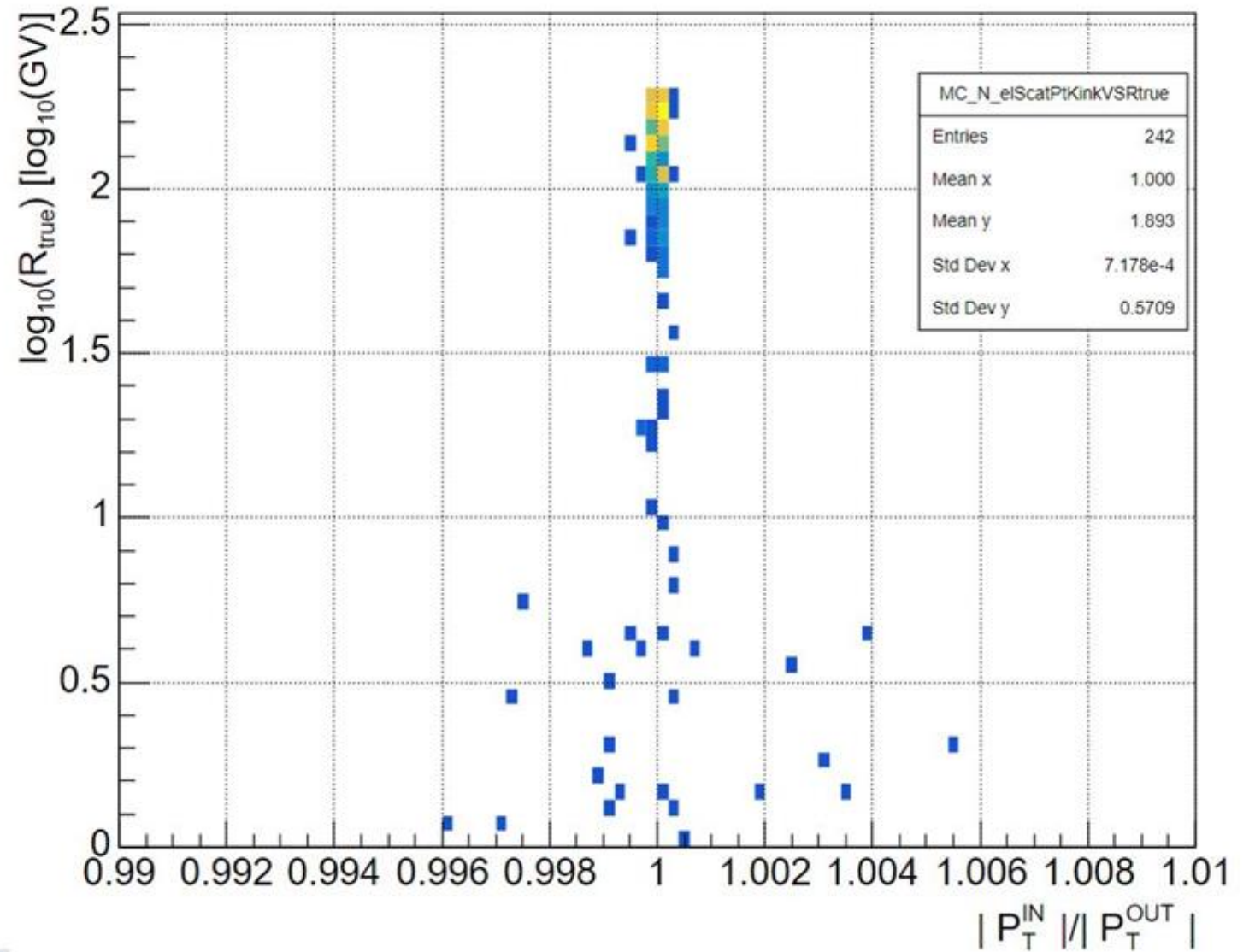
$$R_{gen} \in [150, 200[$$

$$\log_{10}(R_{gen}) < 2.30$$

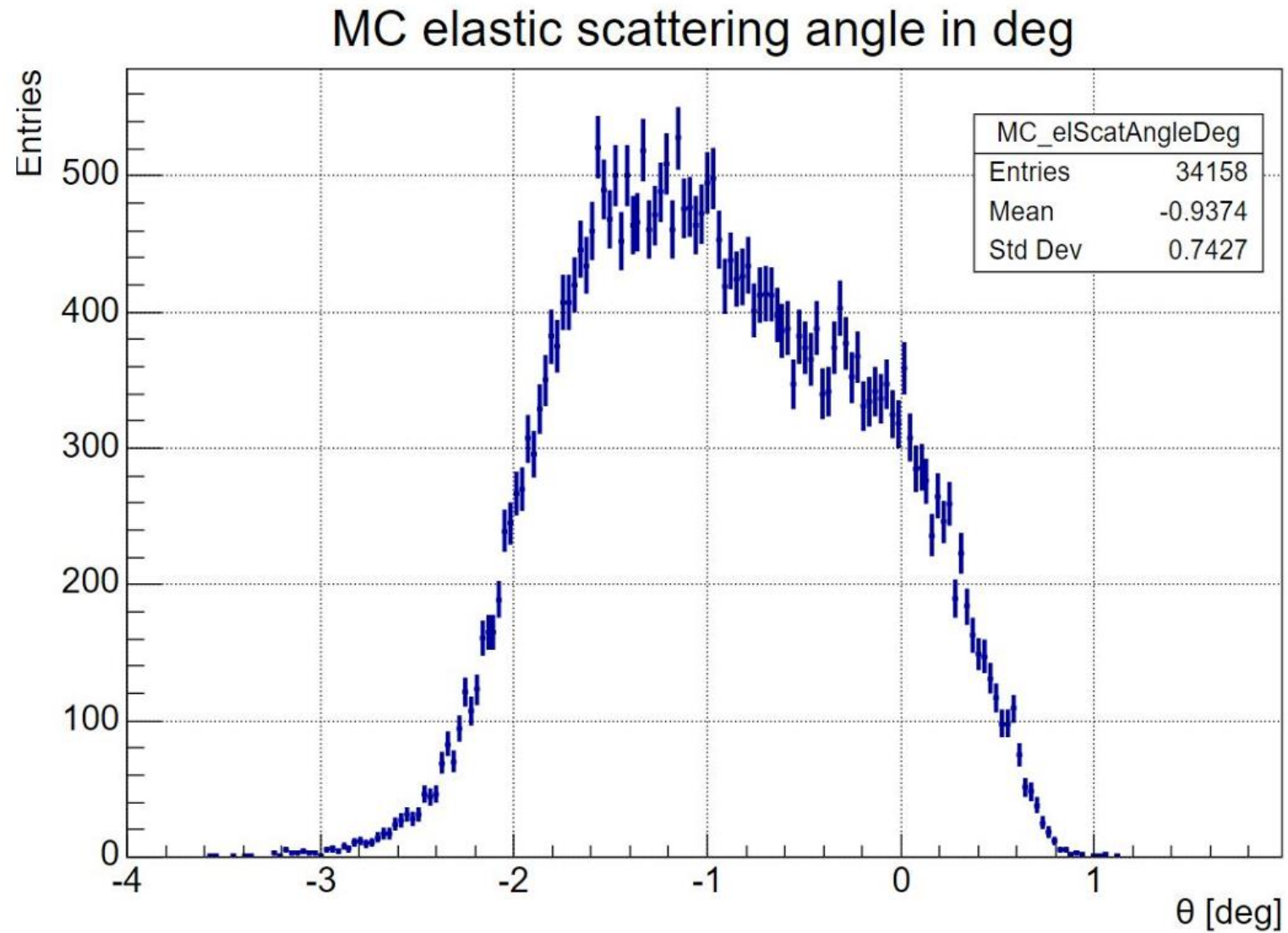


$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

MC elastic scattering momentum variation in module

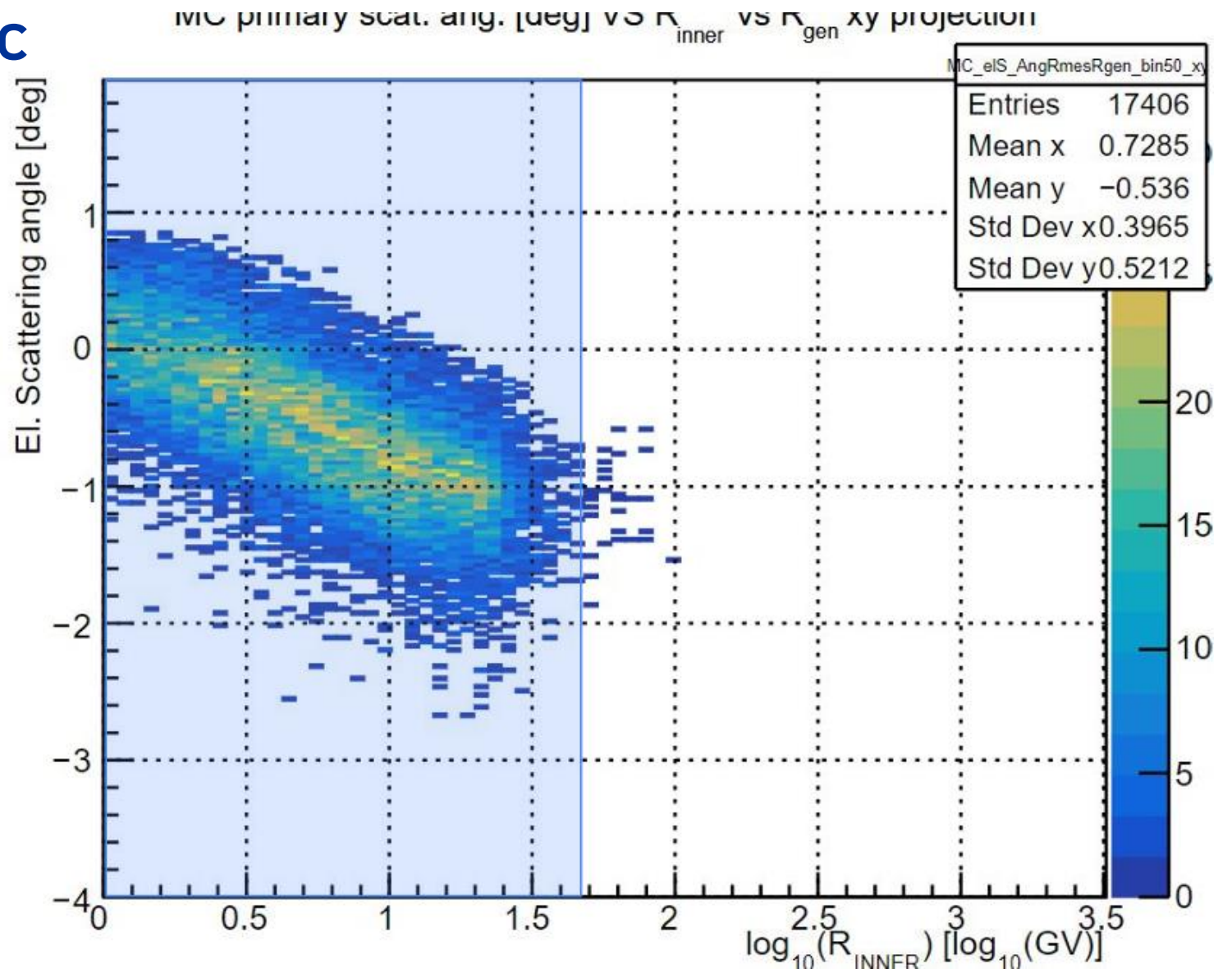


Distribution of the scattering angle ($R > 0$)



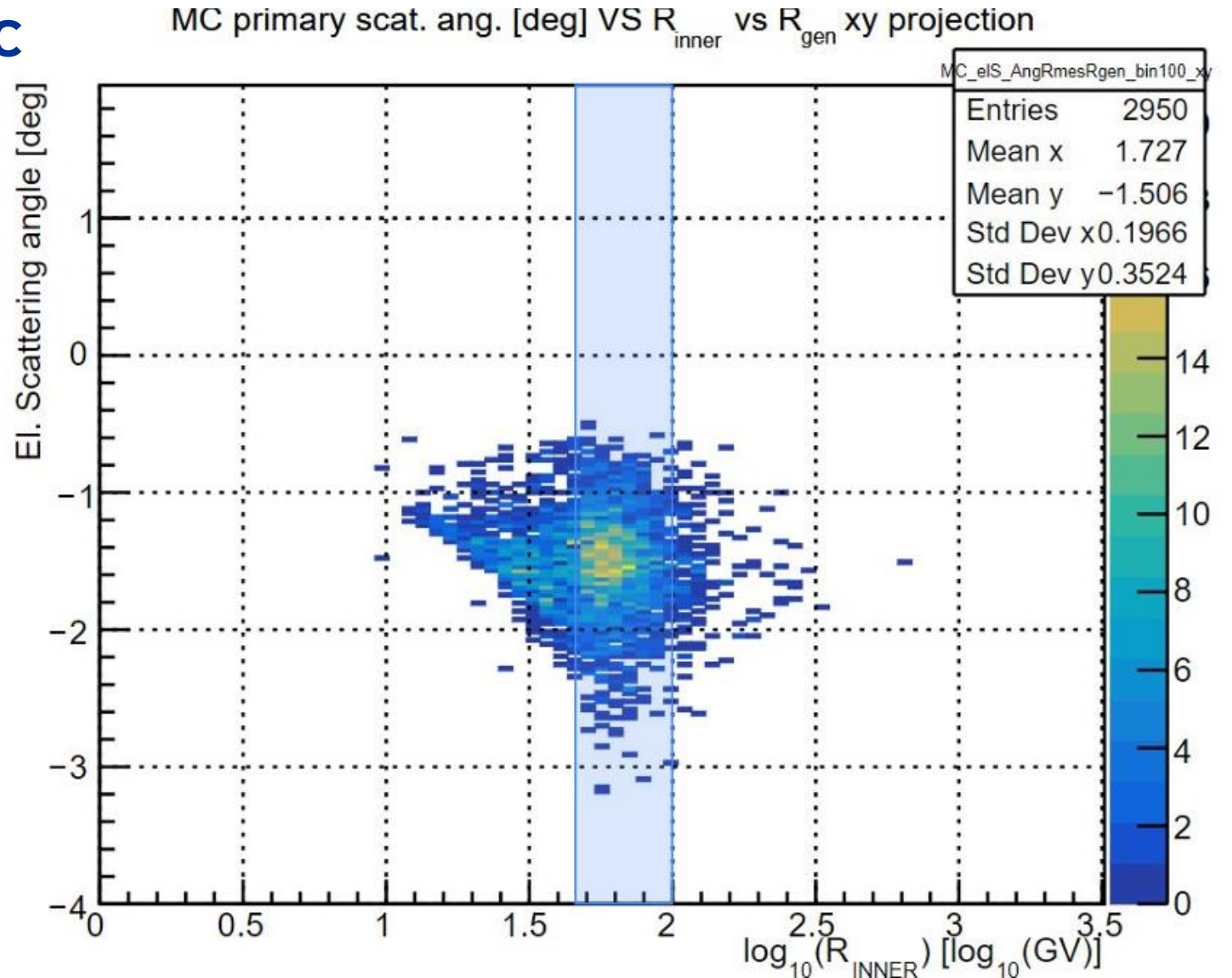
$(R_{inner} > 0)$ and elastic scattering inside the inner tracker

$$R_{gen} \in [50, 100[$$
$$\log_{10}(R_{gen}) < 2.0$$



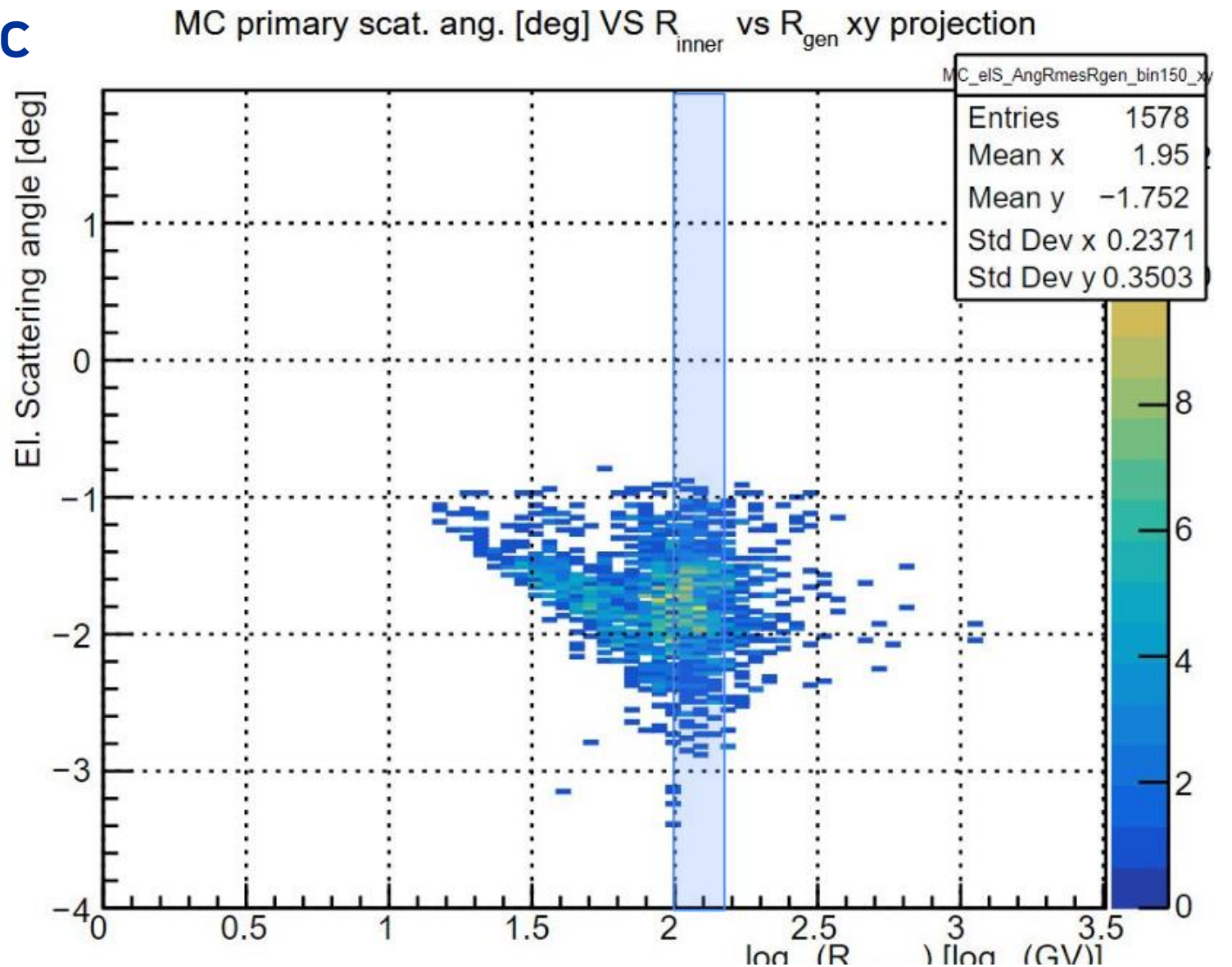
$(R_{inner} > 0)$ and elastic scattering inside the inner tracker

$$R_{gen} \in [50, 100[$$
$$\log_{10}(R_{gen}) < 2.0$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

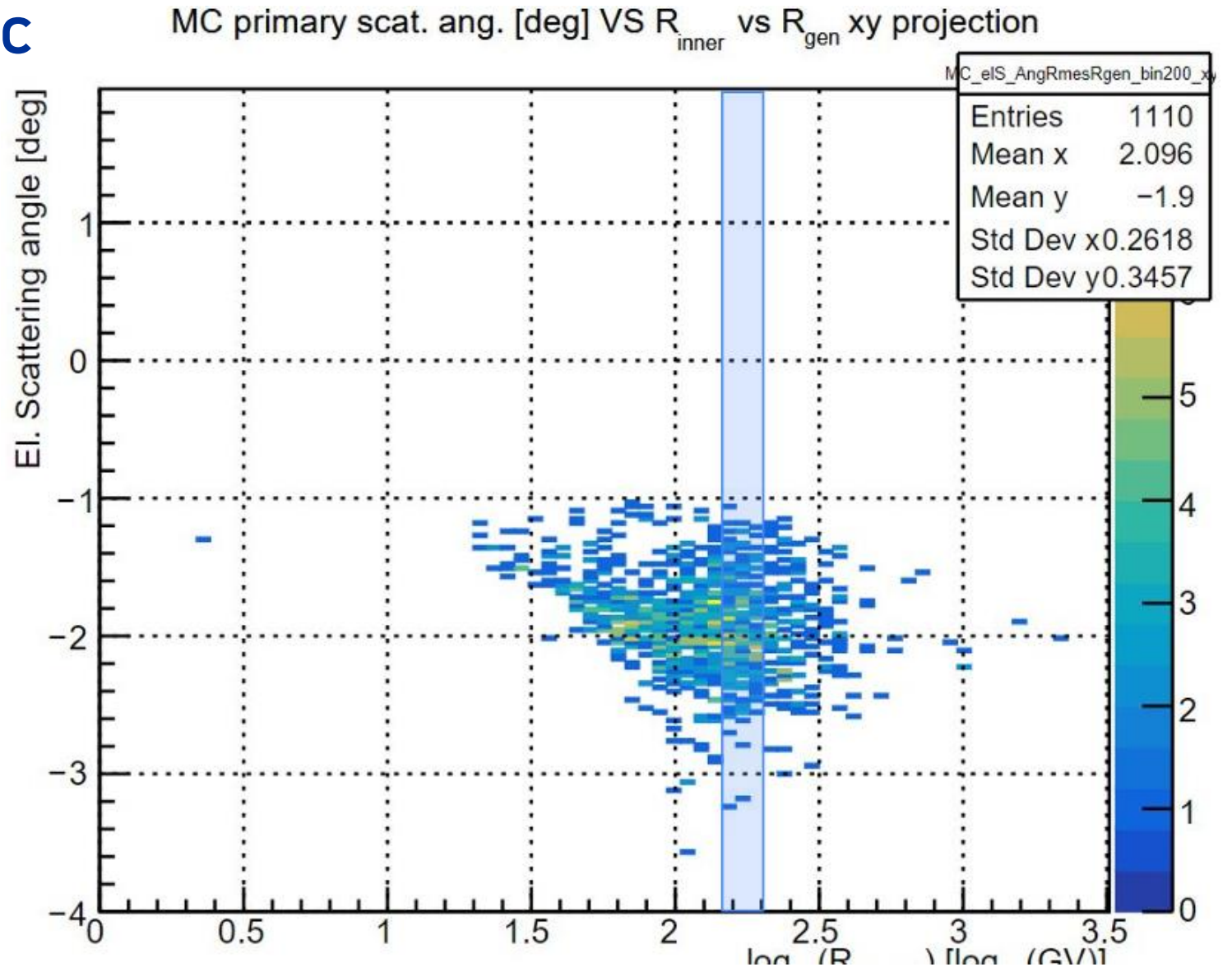
$$R_{gen} \in [100, 150[$$
$$\log_{10}(R_{gen}) < 2.18$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

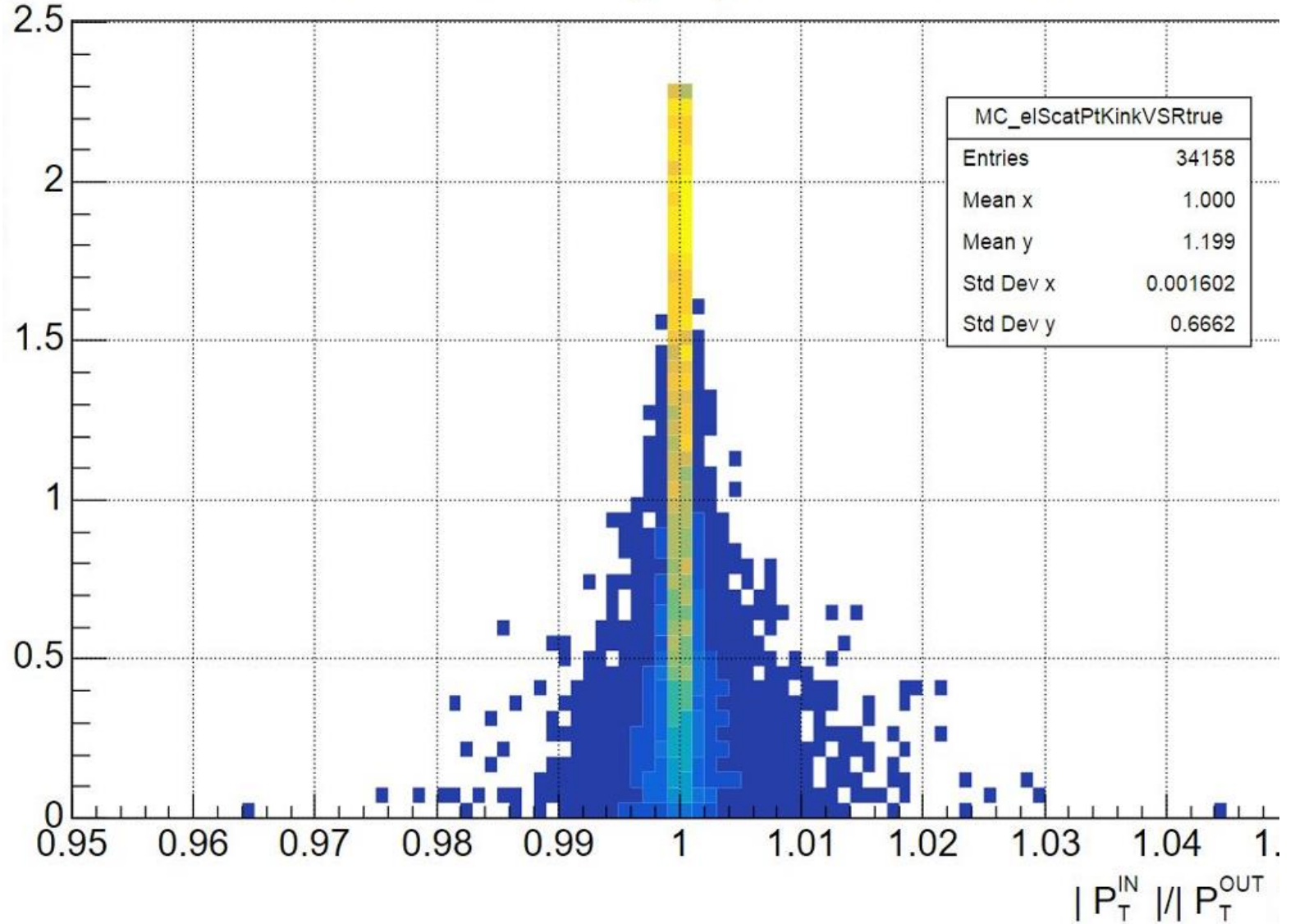
$$R_{gen} \in [150, 200[$$

$$\log_{10}(R_{gen}) < 2.30$$



$(R_{inner} < 0)$ and elastic scattering inside the inner tracker

MC elastic scattering P_T variation in module



Data-Monte Carlo

A Fully Connected Neural Network (FCNN) classifier to characterise the *He* background

- Use the classes previously defined as labels for supervised training
- The Monte Carlo has been weighted using published 7.5 *He* flux.
- Sample composition (**14 % Spillover, 2.5% El. Scat., 53% Had. Inel., 30% Other**)
- **Training** sample ($1.78 \cdot 10^5$) and **validation** sample ($0.76 \cdot 10^5$) events
- Choose variables with good data-MC agreement as input features:

Time Of Flight:

- TOF on-time clusters ($\times 4$)

Inner Tracker:

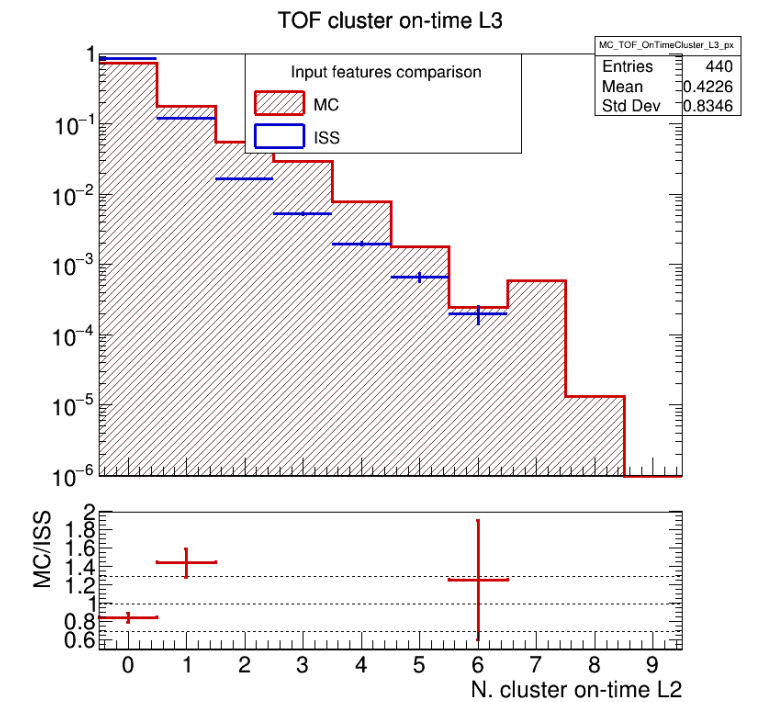
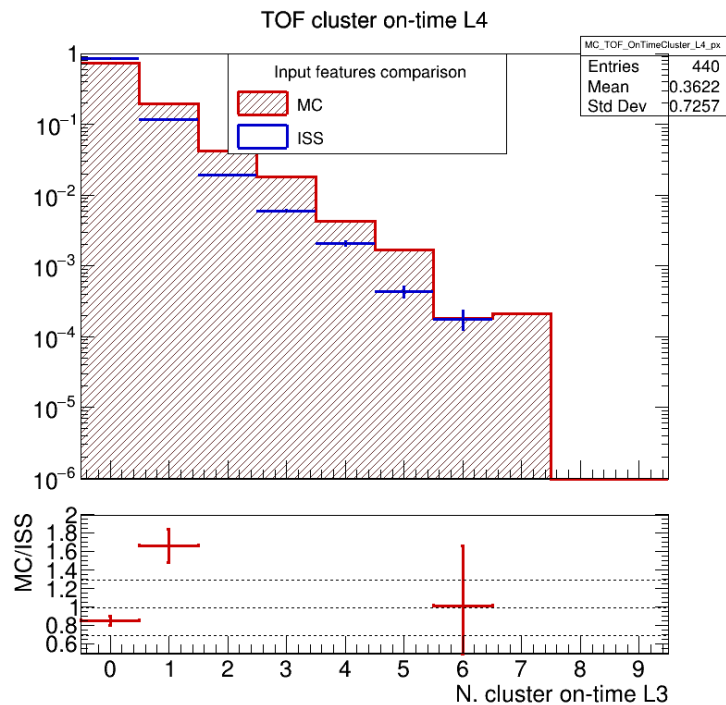
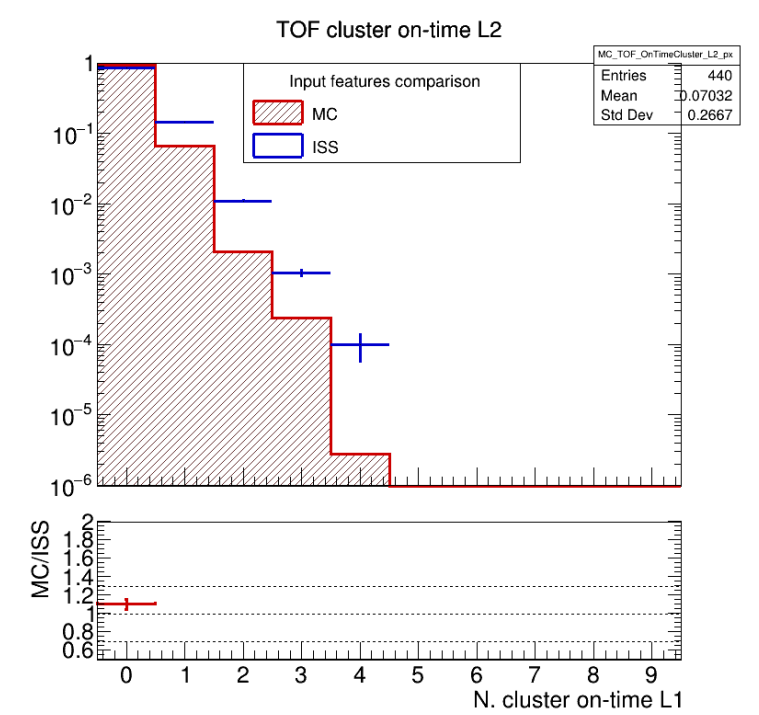
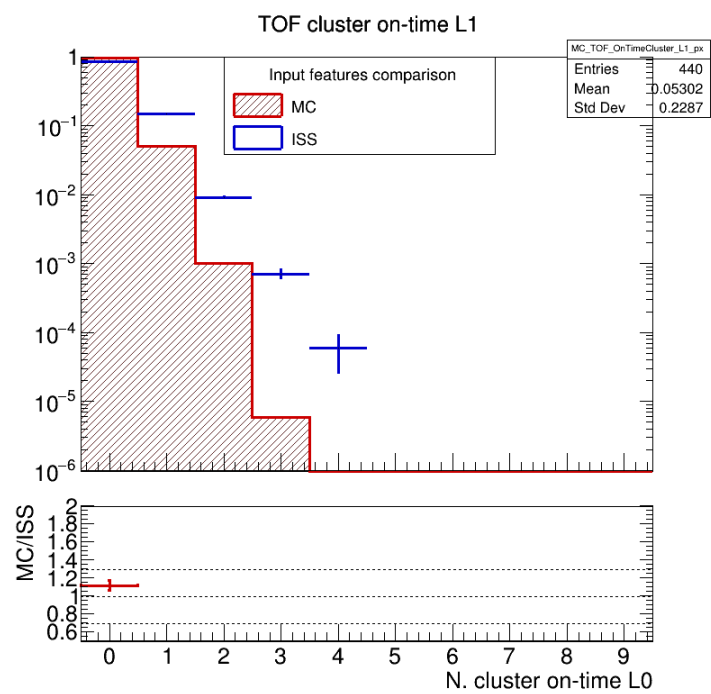
- TRK min feet-distance ($\times 4$)
- TRK Y cluster ($\times 7$)
- TRK Max cluster distance ($\times 7$)
- TRK track hit |Y| ($\times 7$)
- TRK NormEdep2Y* ($\times 7$)

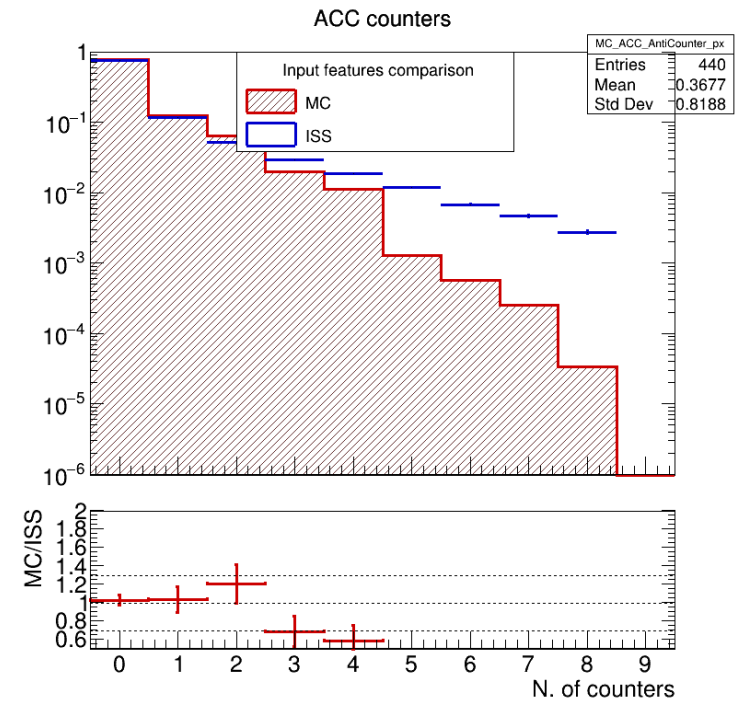
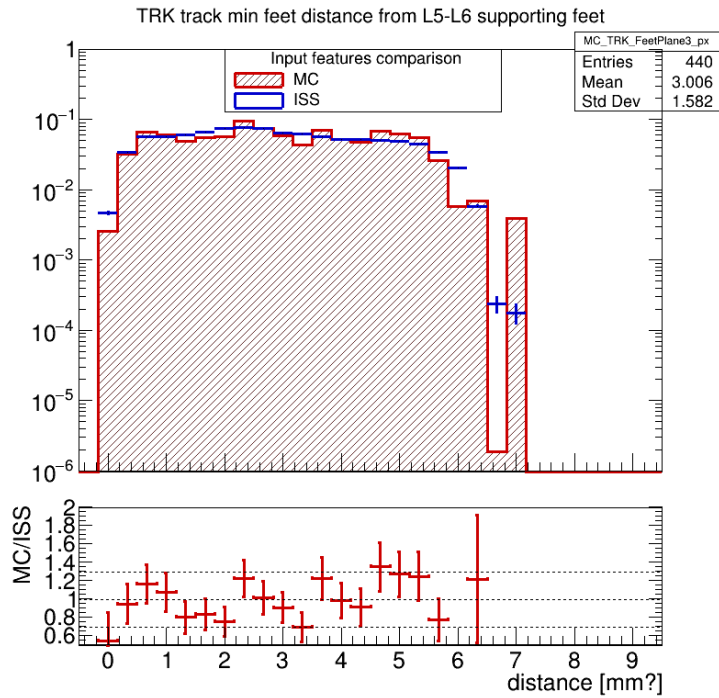
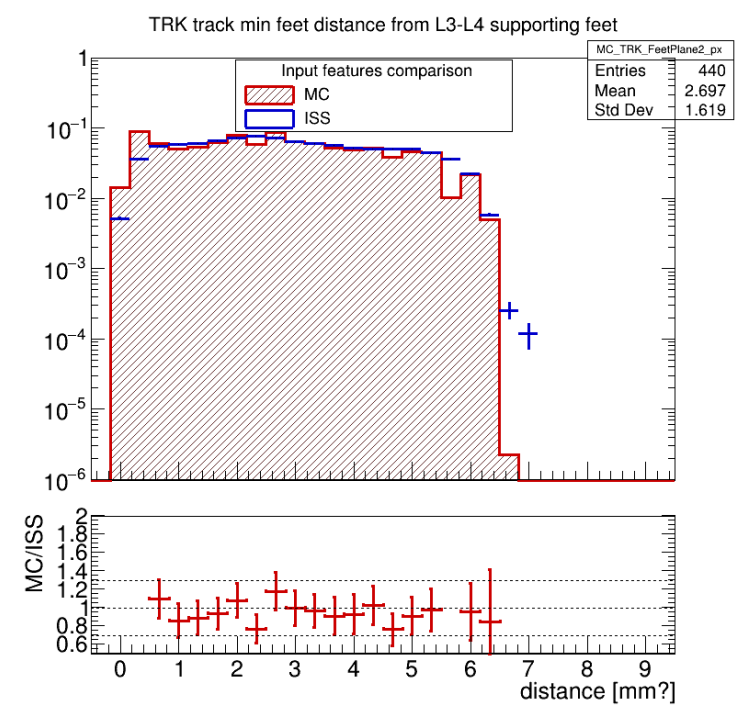
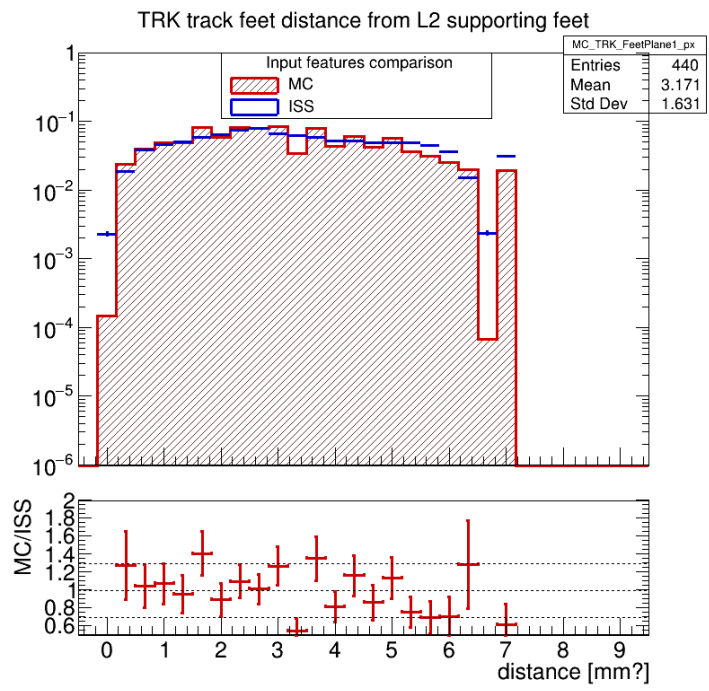
Anti-Coincidence system:

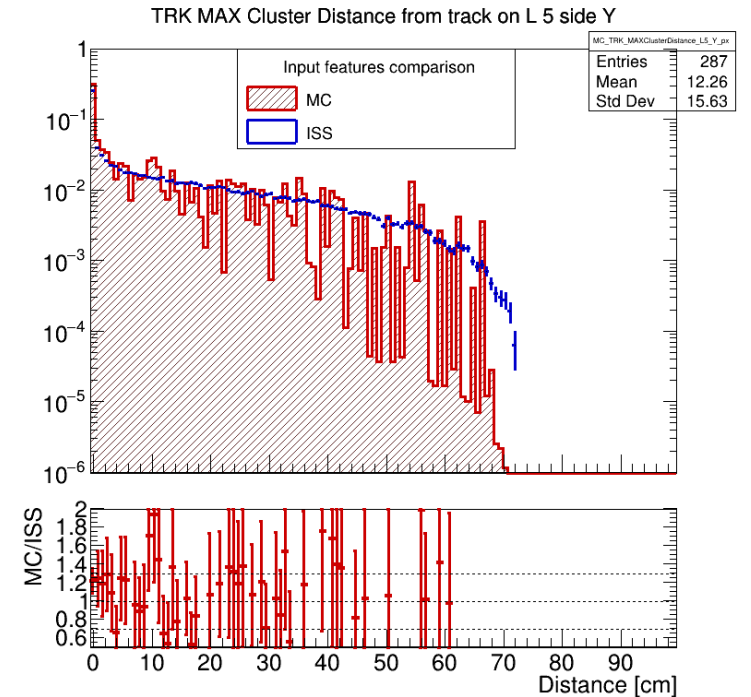
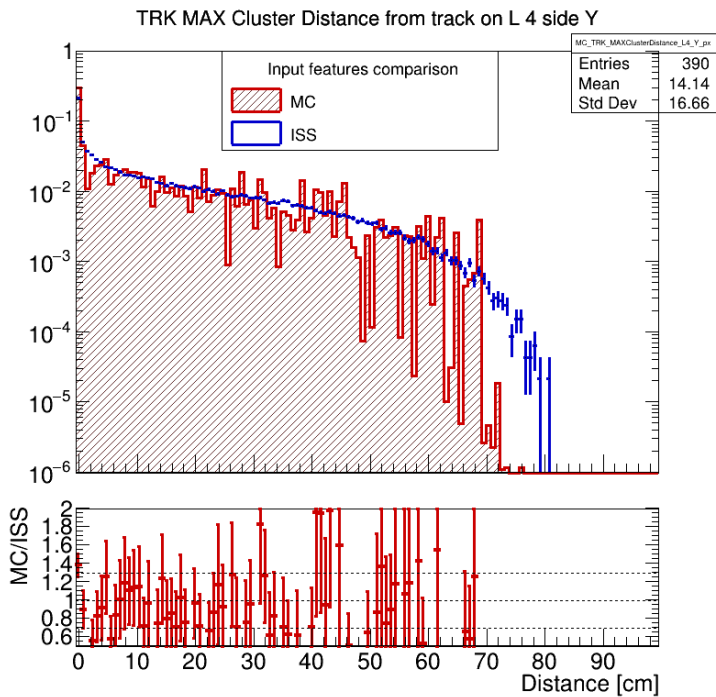
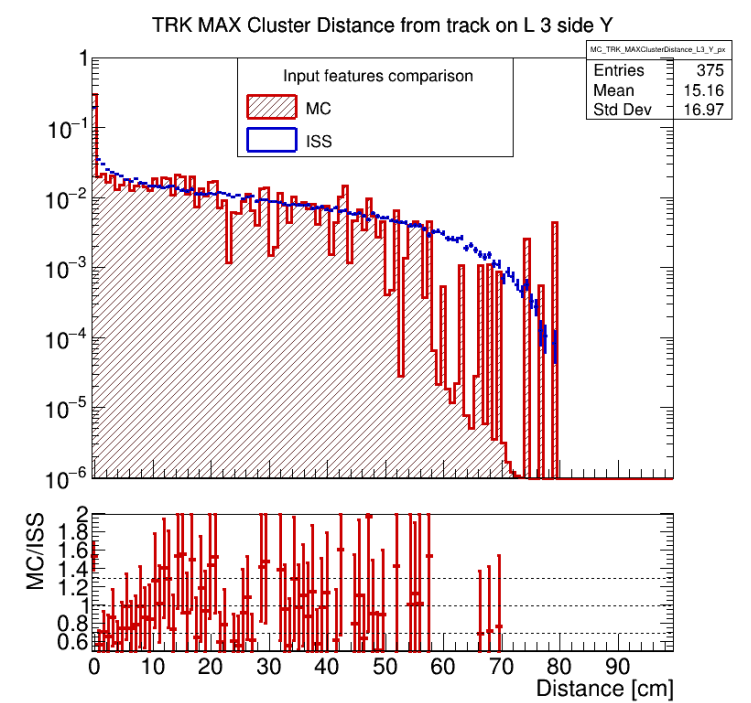
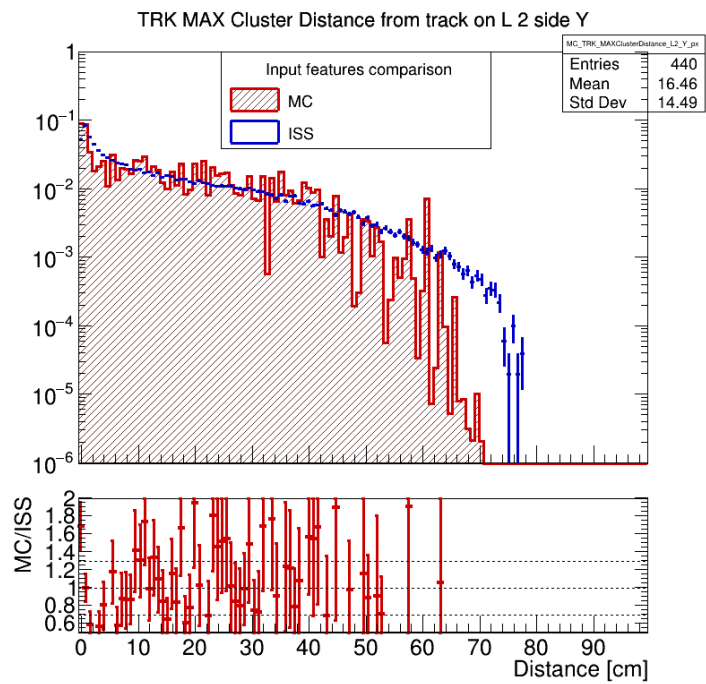
- ACC counters

$$*NormEdep2Y = \frac{Track E_{dep} Y}{Cluster E_{dep} Y(2\text{ cm from Track hit}) + Track E_{dep} Y}$$

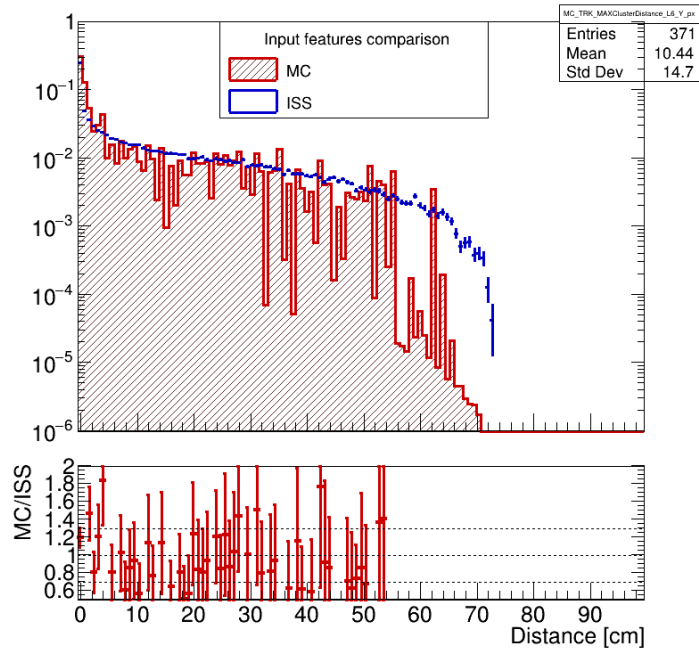
Total number of input features = 37



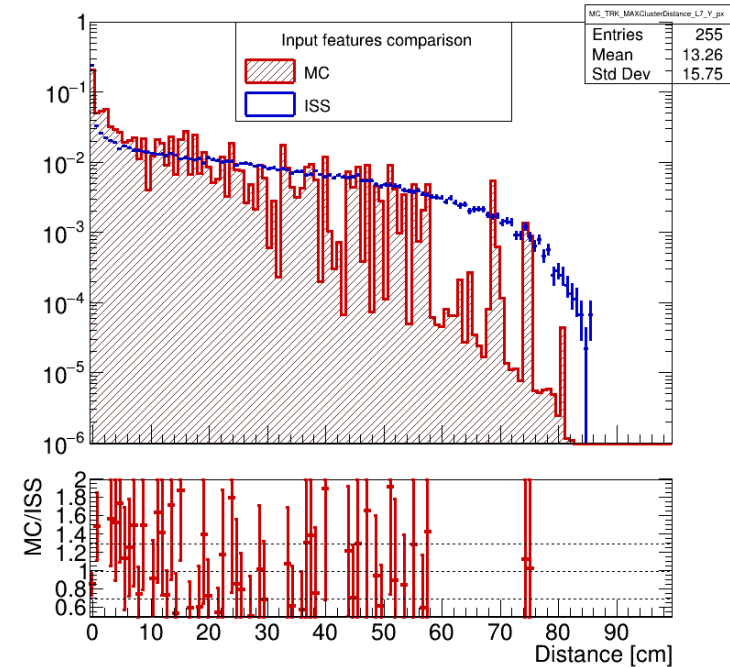




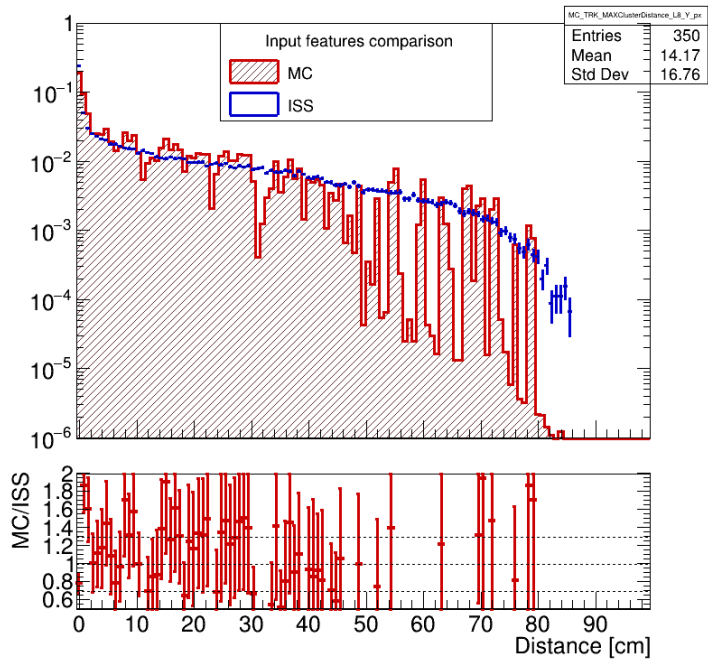
TRK MAX Cluster Distance from track on L 6 side Y

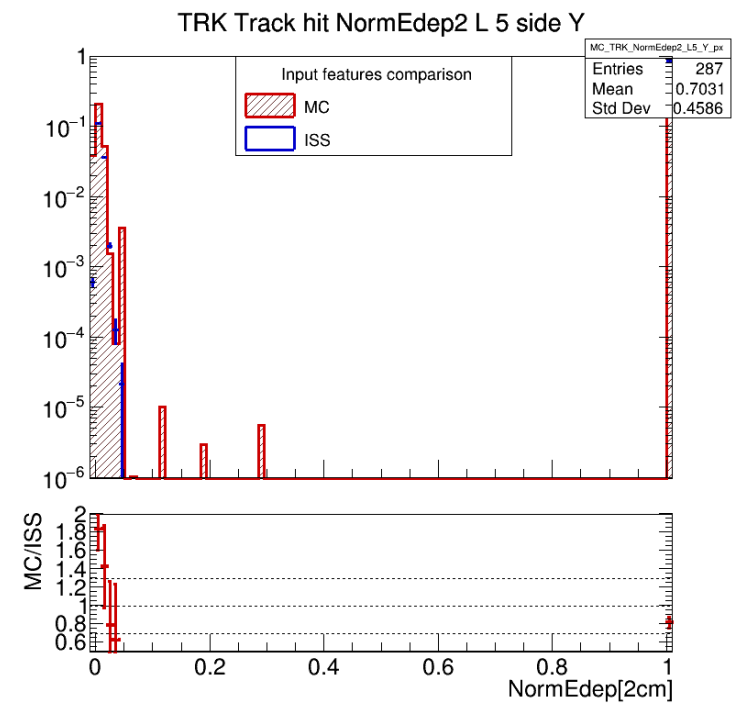
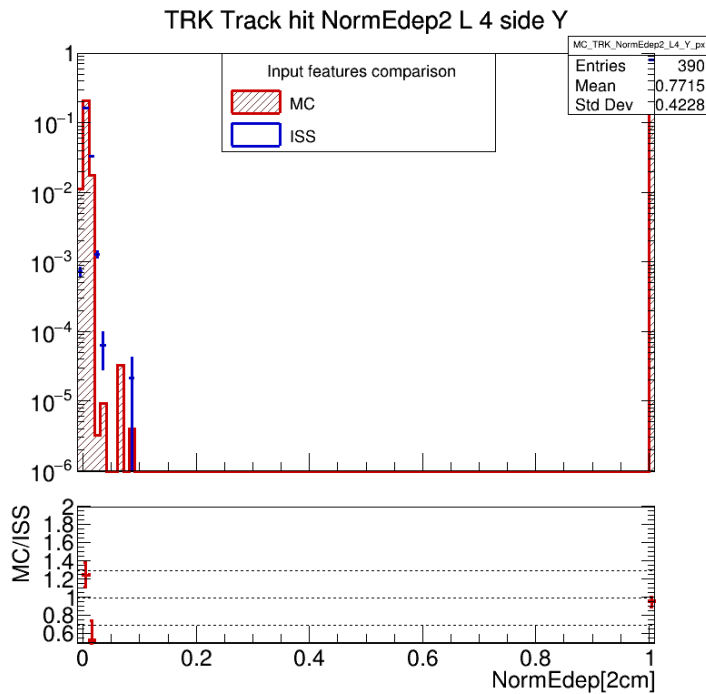
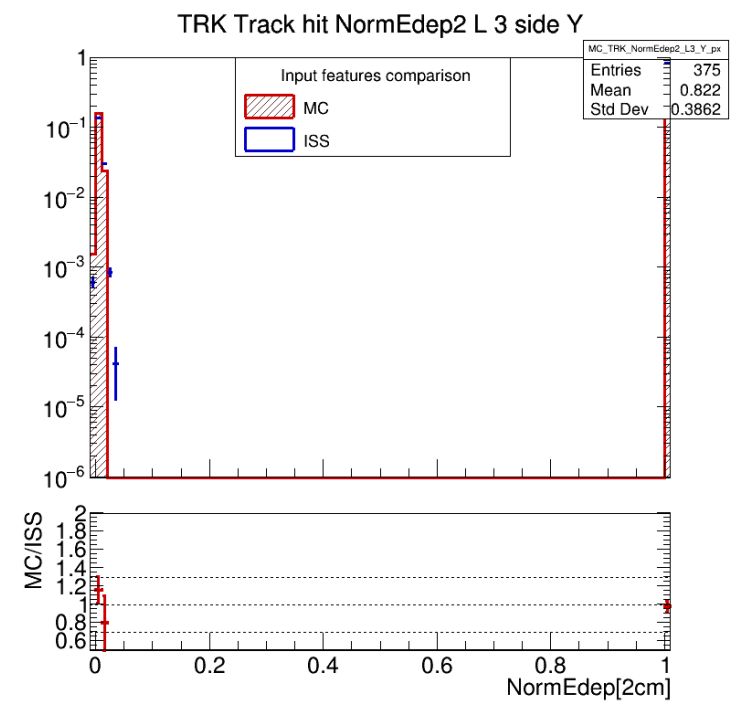
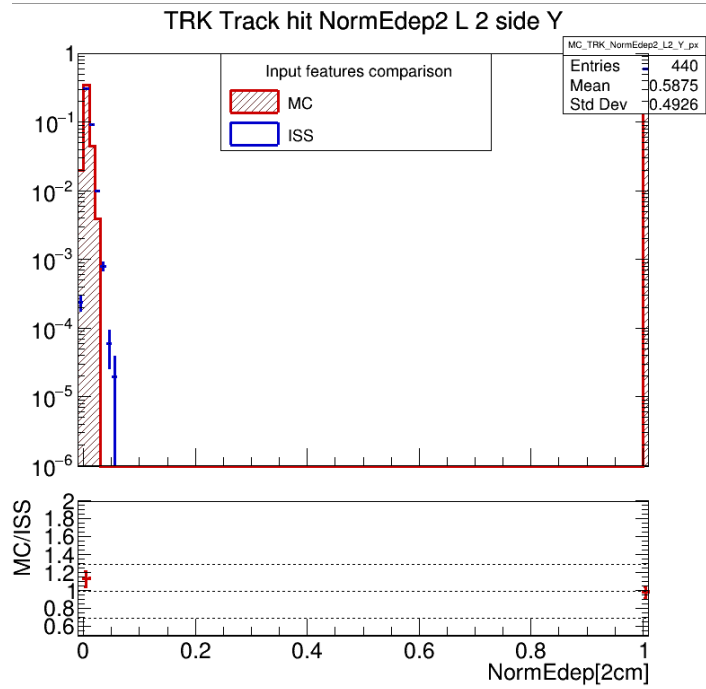


TRK MAX Cluster Distance from track on L 7 side Y

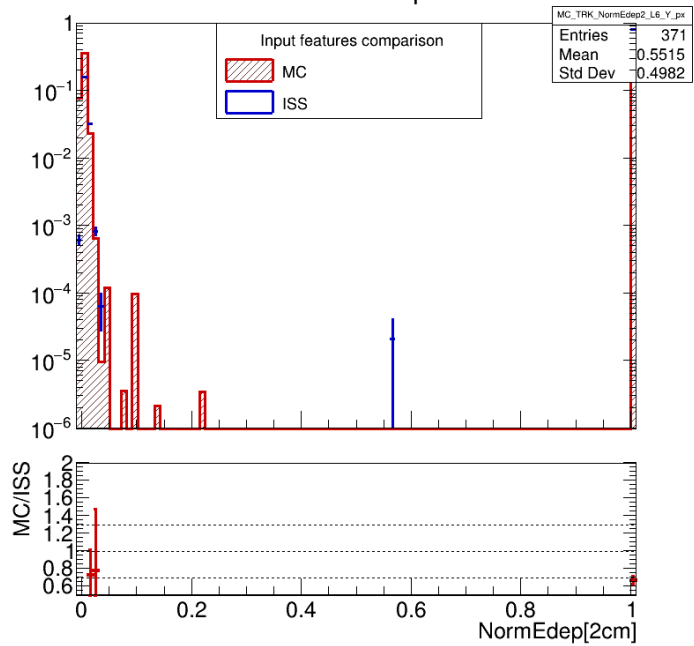


TRK MAX Cluster Distance from track on L 8 side Y

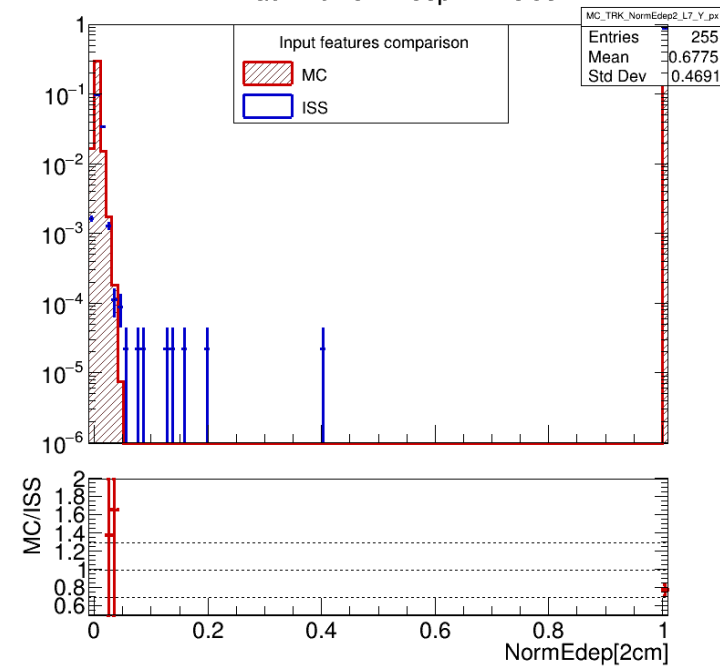




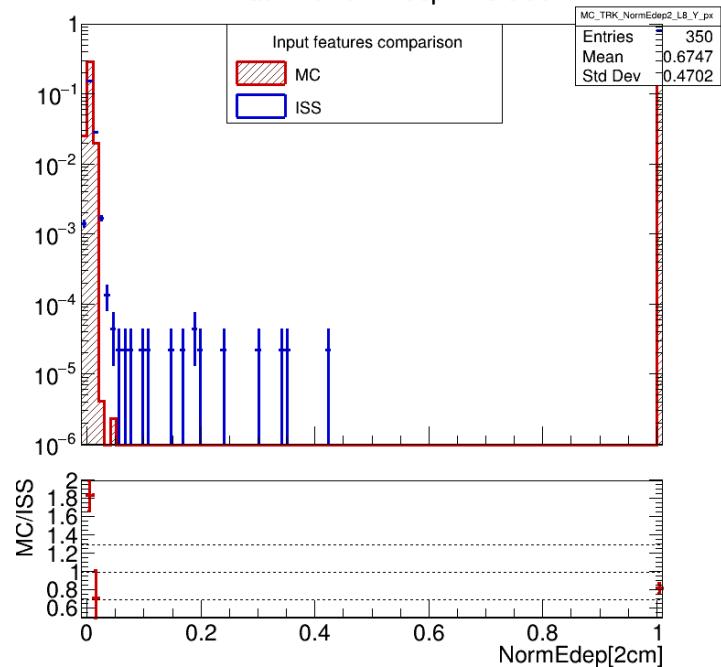
TRK Track hit NormEdep2 L 6 side Y

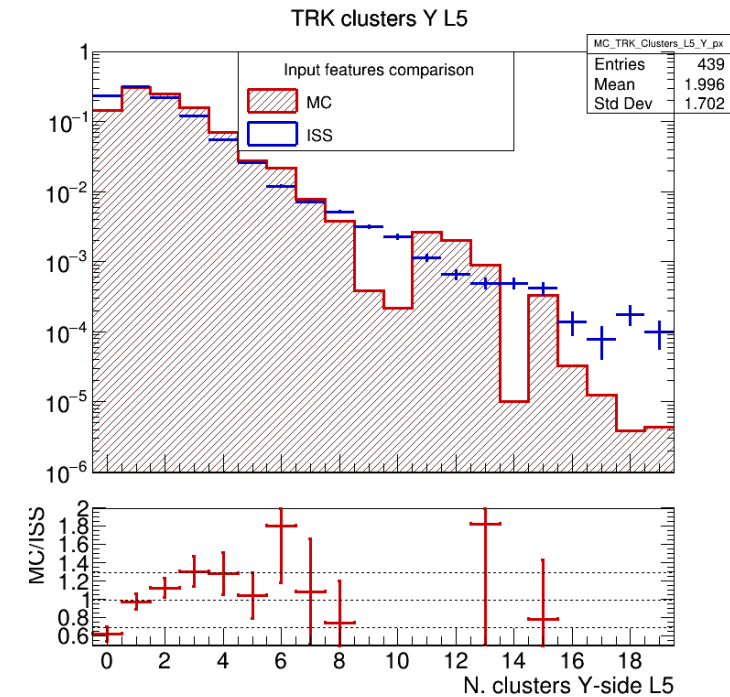
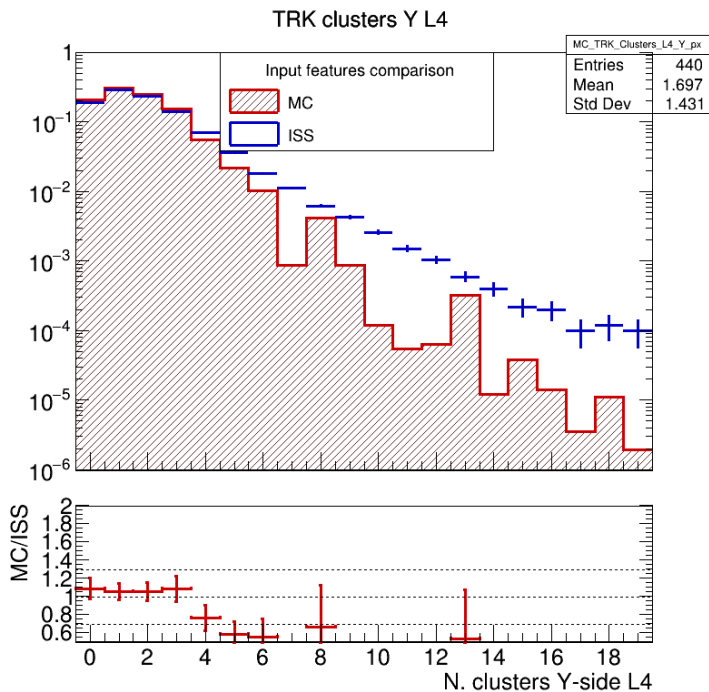
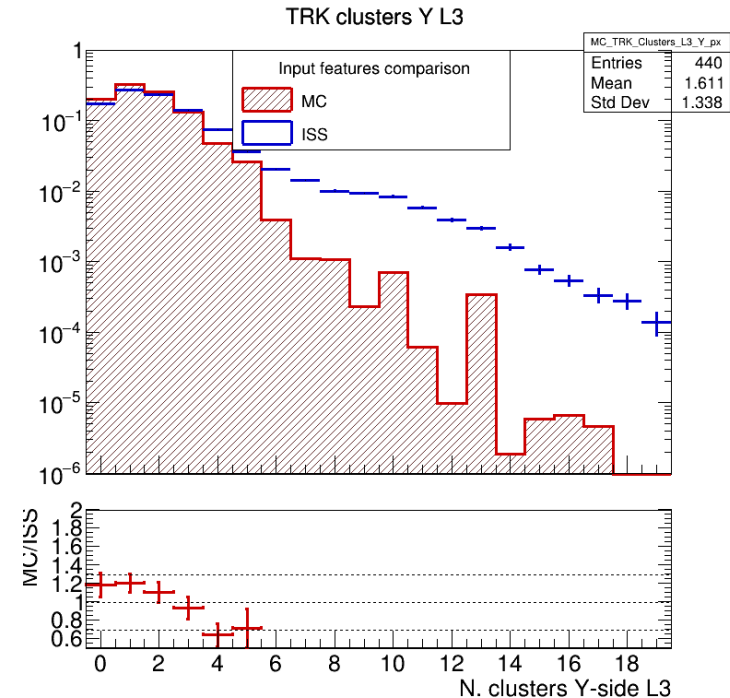
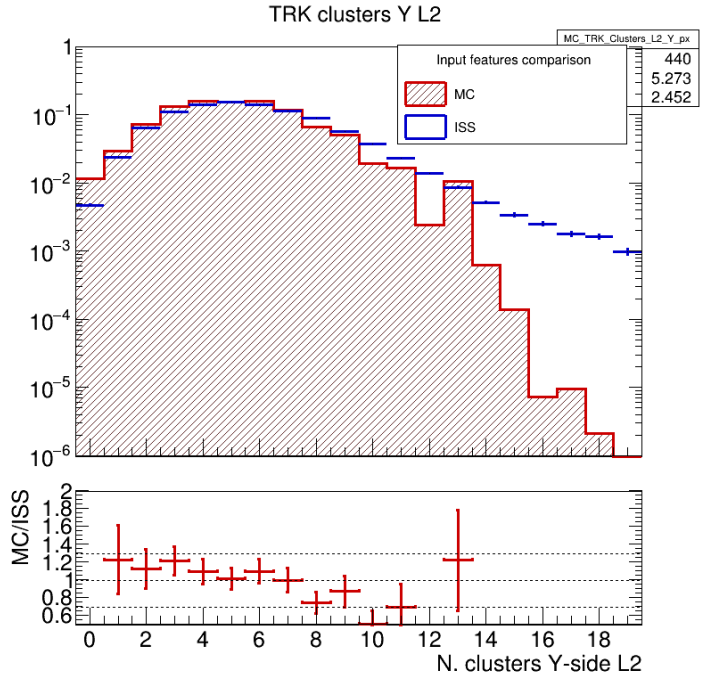


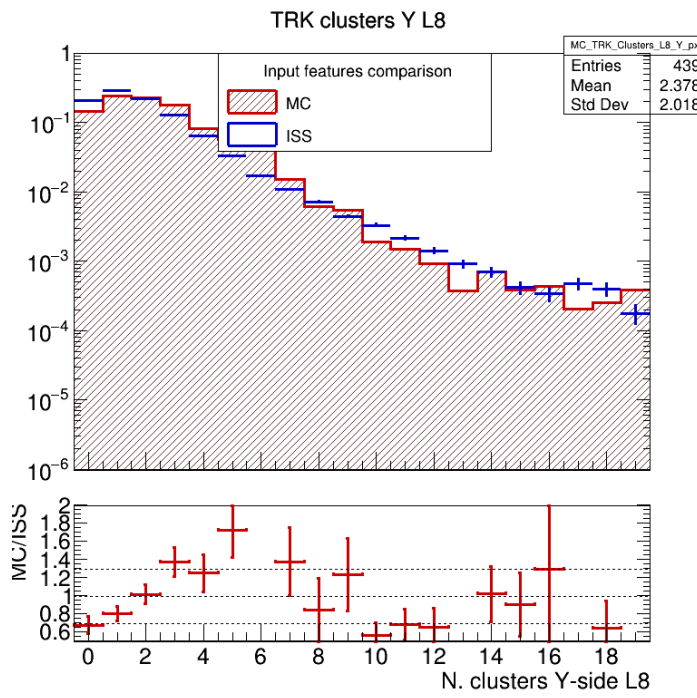
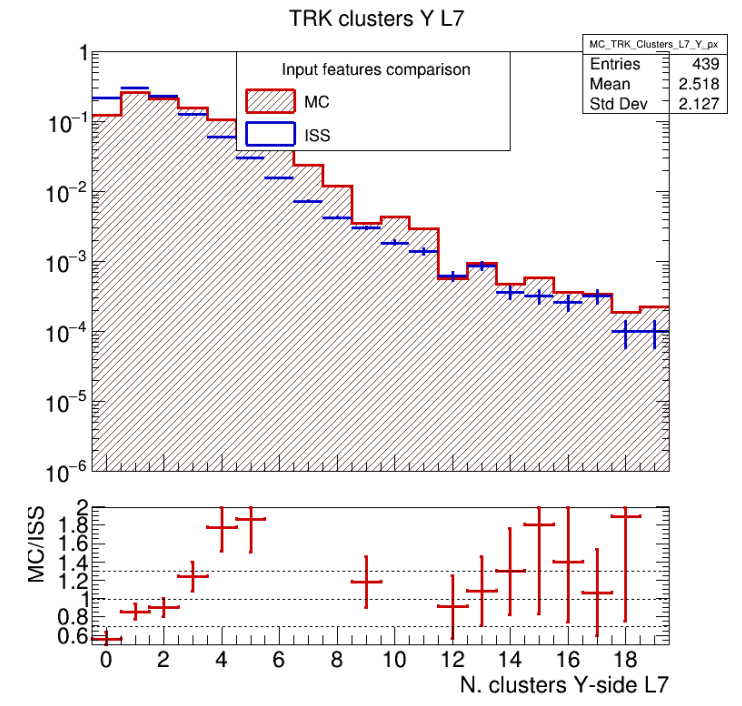
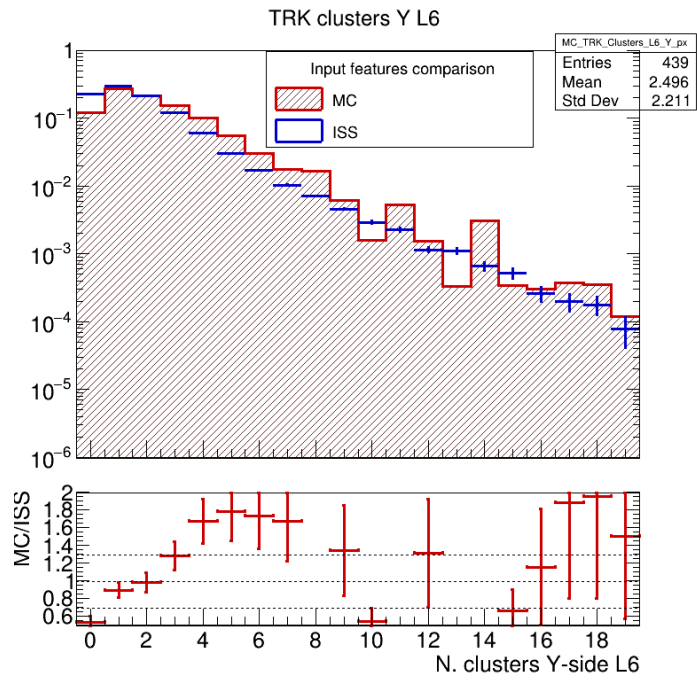
TRK Track hit NormEdep2 L 7 side Y



TRK Track hit NormEdep2 L 8 side Y







Input features