Progress report on machine learning techniques for $\overline{He}$ search in cosmic rays

Francesco Rossi and P. Zuccon

University of Trento

NAIA v1.2.0

Event selection (1):

RTI

Good RTI
Outside SAA
Livetime > 0.5
Zenith $< 40\,°$

TRIGGER

Physical trigger

TOF β

$N.\beta$ clusters ≥ 4
$\beta > 0.3$
$\chi^2_{TIME} < 10$

TRACK

N. tracks ≤ 2
L2 XY hit
Y pattern: ≥ 5 between L3-L8
X pattern: ≥ 4, L3||L4, L5||L6, L7||L8

TRACKER

Z Inner Tracker $\in [1.7, 2.4]$
Inner fiducial volume
χ^2_Y (inner) < 10
χ^2_X (inner) < 10

Event selection (2):

TOF charge

Q upper TOF $\in [1.5, 3.0]$
Q lower TOF ≥ 1.5

RIGIDITY

R_{INNER} signs coherence
$|R_{INNER}| > 1.2$ GV

Inner Upper Half = R_{UH}
Inner Lower Half = R_{LH}
Inner span rig. = R_{INNER}

$(\boldsymbol{R_{inner}} > \mathbf{0}) \rightarrow (\boldsymbol{R_{UH}} > \mathbf{0}, \boldsymbol{R_{LH}} > \mathbf{0})$
$(\boldsymbol{R_{inner}} < \mathbf{0}) \rightarrow (\boldsymbol{R_{UH}} < \mathbf{0}, \boldsymbol{R_{LH}} < \mathbf{0})$

RICH (if present)

$\beta_{RICH} < 1$
$|\beta_{RICH} - \beta_{TOF}| \leq 0.06$
Ring hits ≥ 4

TOF β

IF $\beta_{TOF} \geq 0.96$ require RICH

Total number of events

	$\boldsymbol{R_{INNER}} > 0$	$\boldsymbol{R_{INNER}} < 0$
ISS-data (12.5 y)	$\mathbf{2.7 \times 10^7}$ (prescaled, 1/100)	$\mathbf{0.64 \times 10^5}$
$^4\boldsymbol{He}$ **MC**	$\mathbf{3.7 \times 10^7}$ (prescaled, 1/100)	$\mathbf{6.7 \times 10^5}$
$^3\boldsymbol{He}$ **MC**	$\mathbf{0.07 \times 10^7}$ (prescaled, 1/100)	$\mathbf{0.89 \times 10^5}$

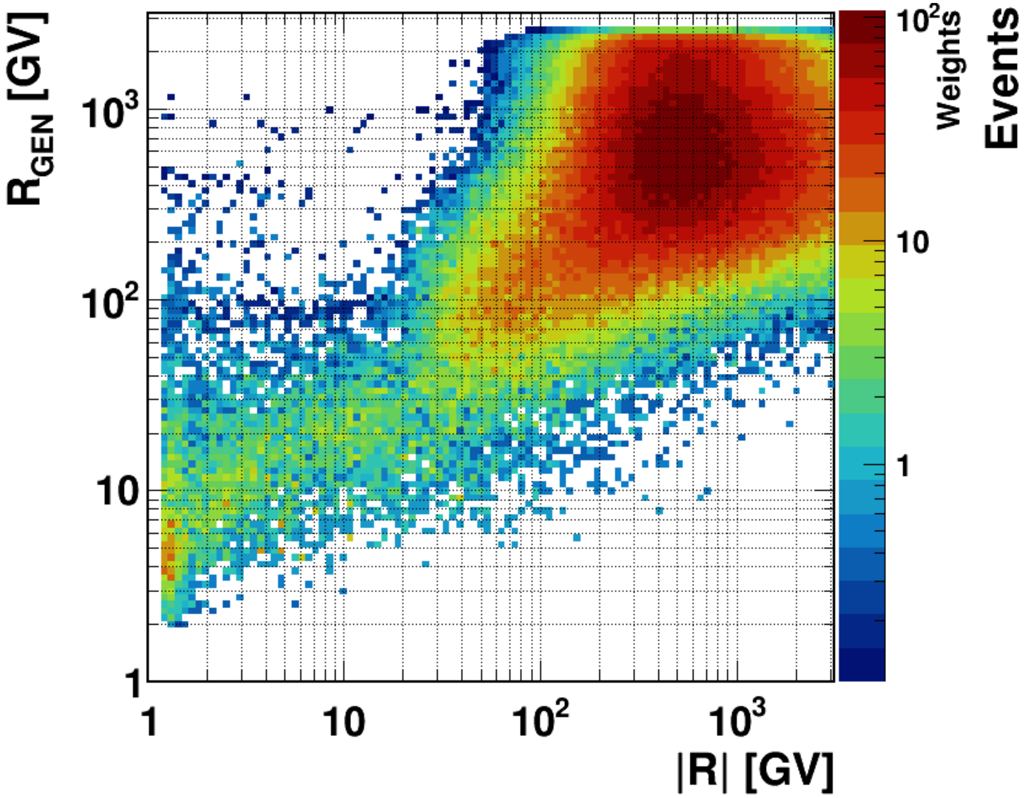
Expected background

Three **charge confusion sources** are known and expected:

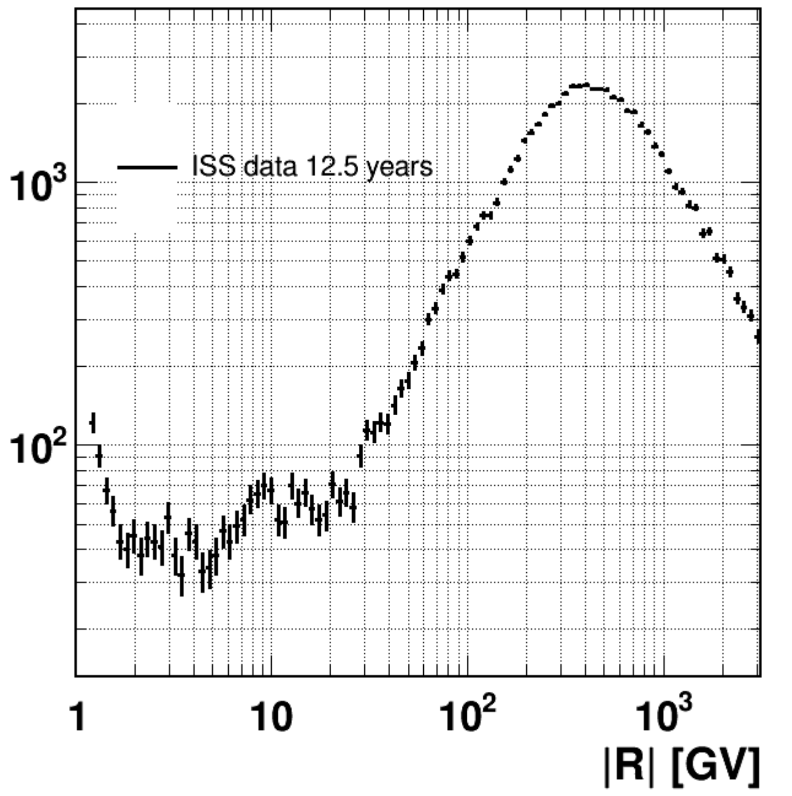
- **Spillover**:
  - Stochastic process, (tracker resolution).
  - Dominant at $R \gtrsim MDR^*$.
- Interactions inside the detector:
  - $^4He \rightarrow {}^3He$ **inelastic interaction**
  - **Large angle scattering**
  - Relevant at $R\ [GV] \in [1, MDR]$

Monte Carlo confusion matrix

MC $^4\mathrm{He}$, $^3\mathrm{He}$.B1315 $(R_{INNER} < 0)$



ISS-data, without geomagnetic cutoff $(R_{INNER} < 0)$



*Maximum Detectable Rigidity

Analysis strategy on ${}^{4}He$, ${}^{3}He$ Monte Carlo

- **Two networks:** one classifier (CL_{MC}) and one autoencoder (AE_{MC}).
  - **CL_{MC} and AE_{MC} are trained on ${}^{4}He$, ${}^{3}He$ MC (R>0),** tested on MC (R<0).
  - Performances study and debugging.

- **Networks trained on MC are not used on ISS data**: avoiding data-MC disagreement effects.

- **Estimate background contribution** due to charged confused He.

Analysis strategy on ISS data

- **Two networks:** one classifier (CL_{ISS}) and one autoencoder (AE_{ISS}).
  - **CL_{ISS} and AE_{ISS} are trained on ISS data (R>0),** tested on ISS (R<0).

- **Models trained on ISS** are used to **identify possible $\overline{He}$ candidates.**
  - Number of candidates as function of the selection on the network's outputs.

- **The reconstructed mass** of the remaining events **is used as final check**.

Supervised classifier

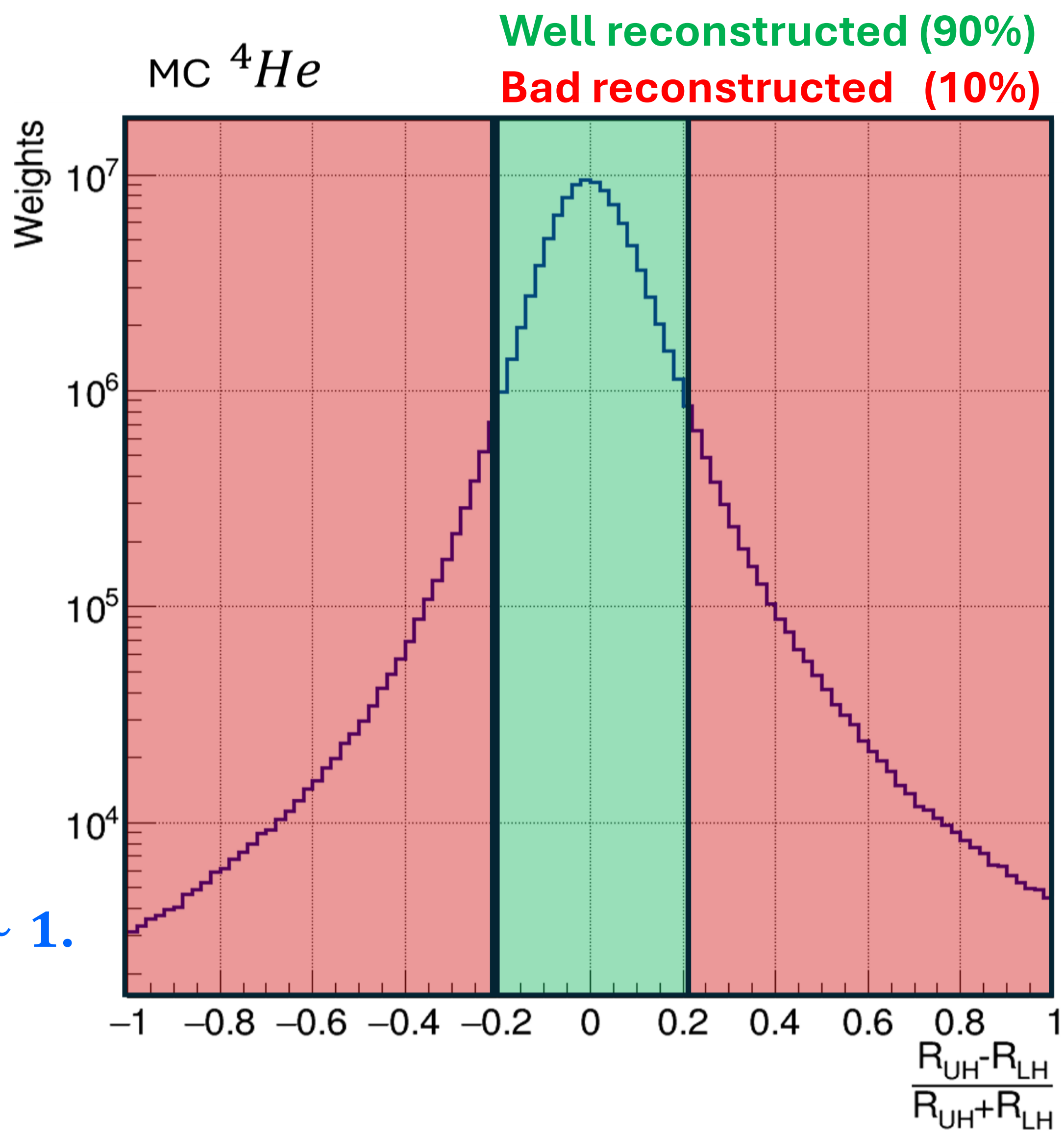
- Goal: **distinguish well-reconstructed rigidities from poorly reconstructed ones**.

- The majority of R < 0 events are poorly reconstructed.

- Define a data-driven label:
  - **Based on*** $\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}$
  - sensitive to spillover.

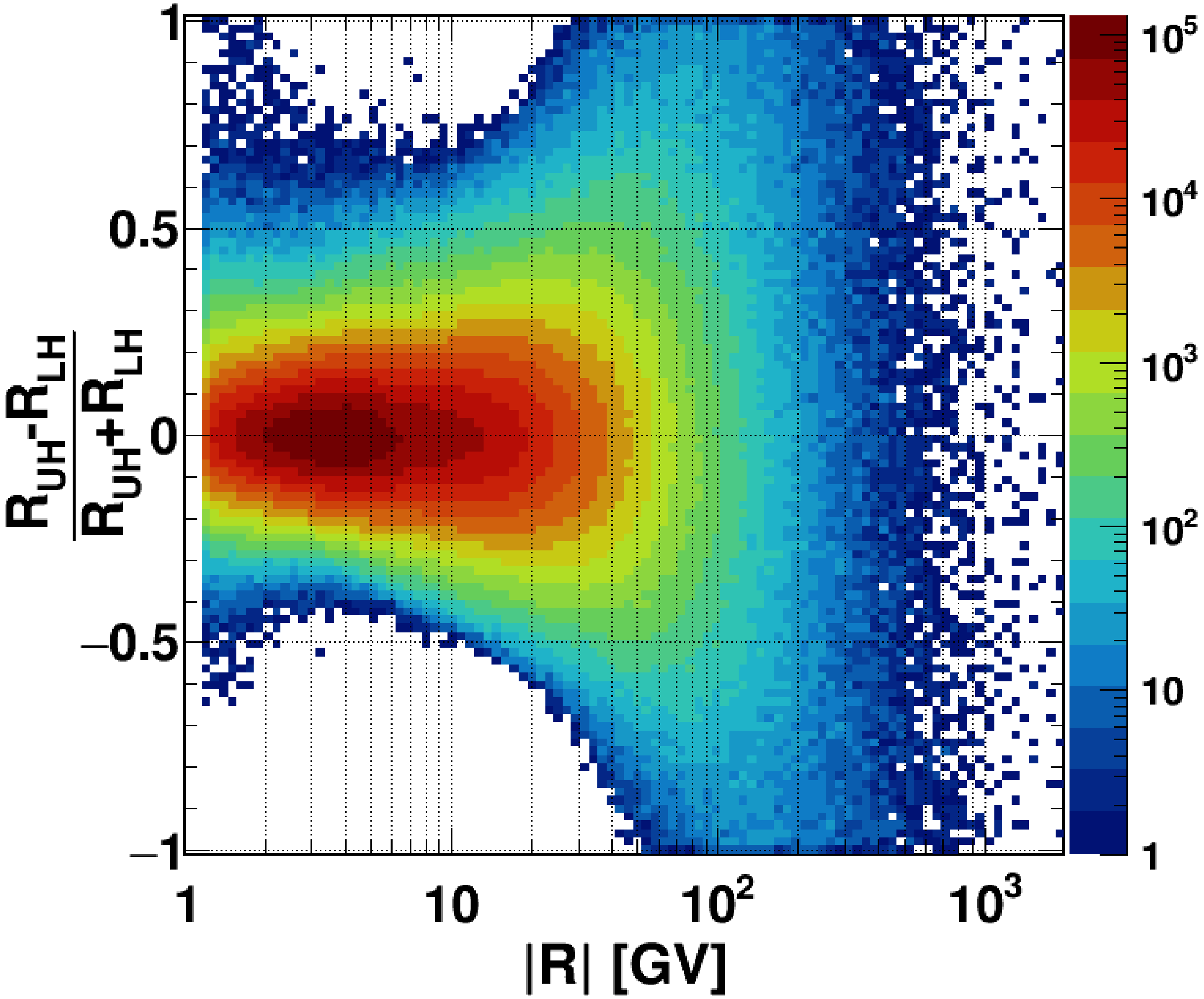
* R_{UH} = Inner Upper Half
  R_{LH} = Inner Lower Half

- **Reliable $\overline{He}$ candidates should have a (CL_{ISS}) score ~ 1.**

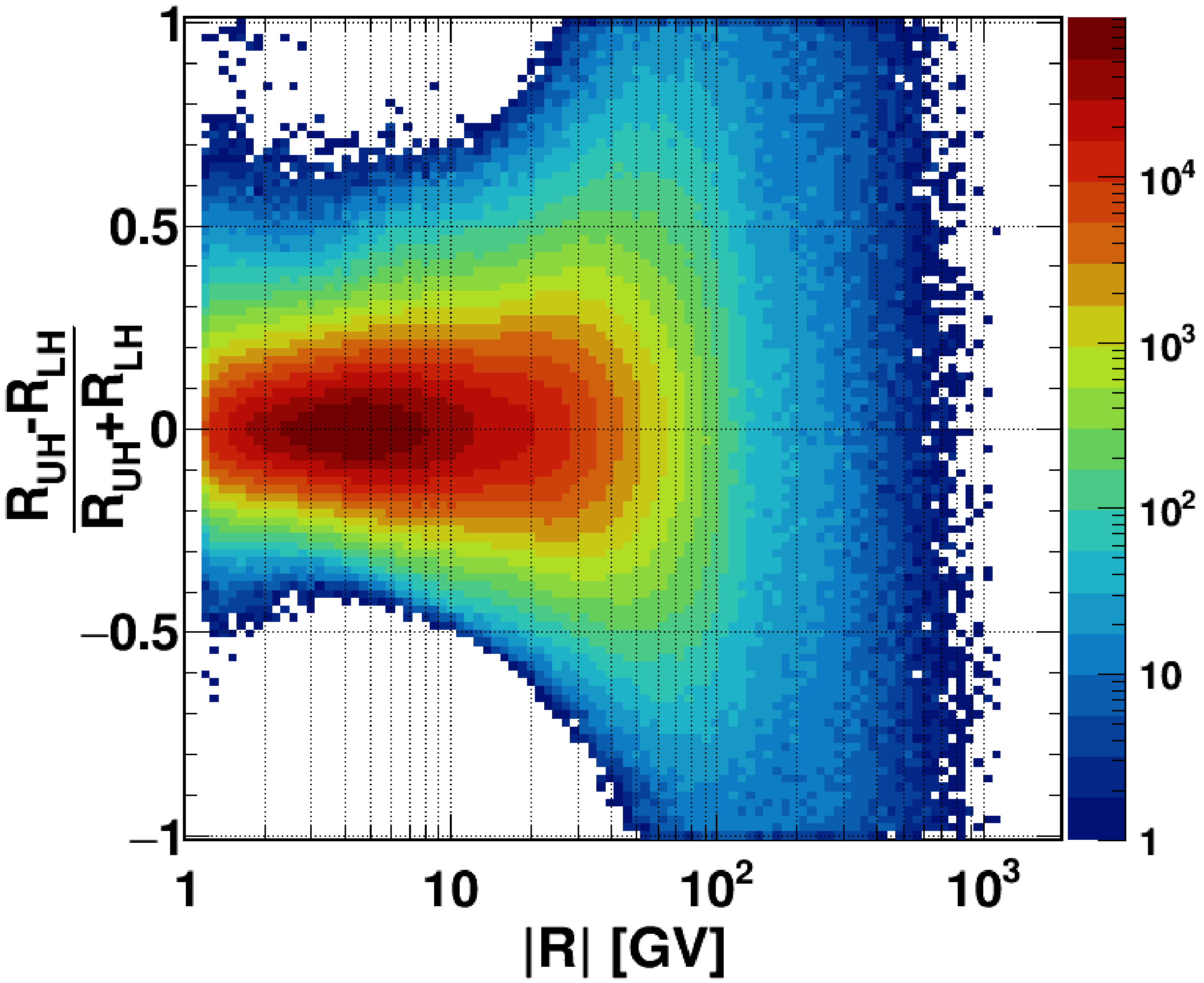


Classifier label dependency from Rigidity (R>0)

ISS data $R_{INNER} > 0$

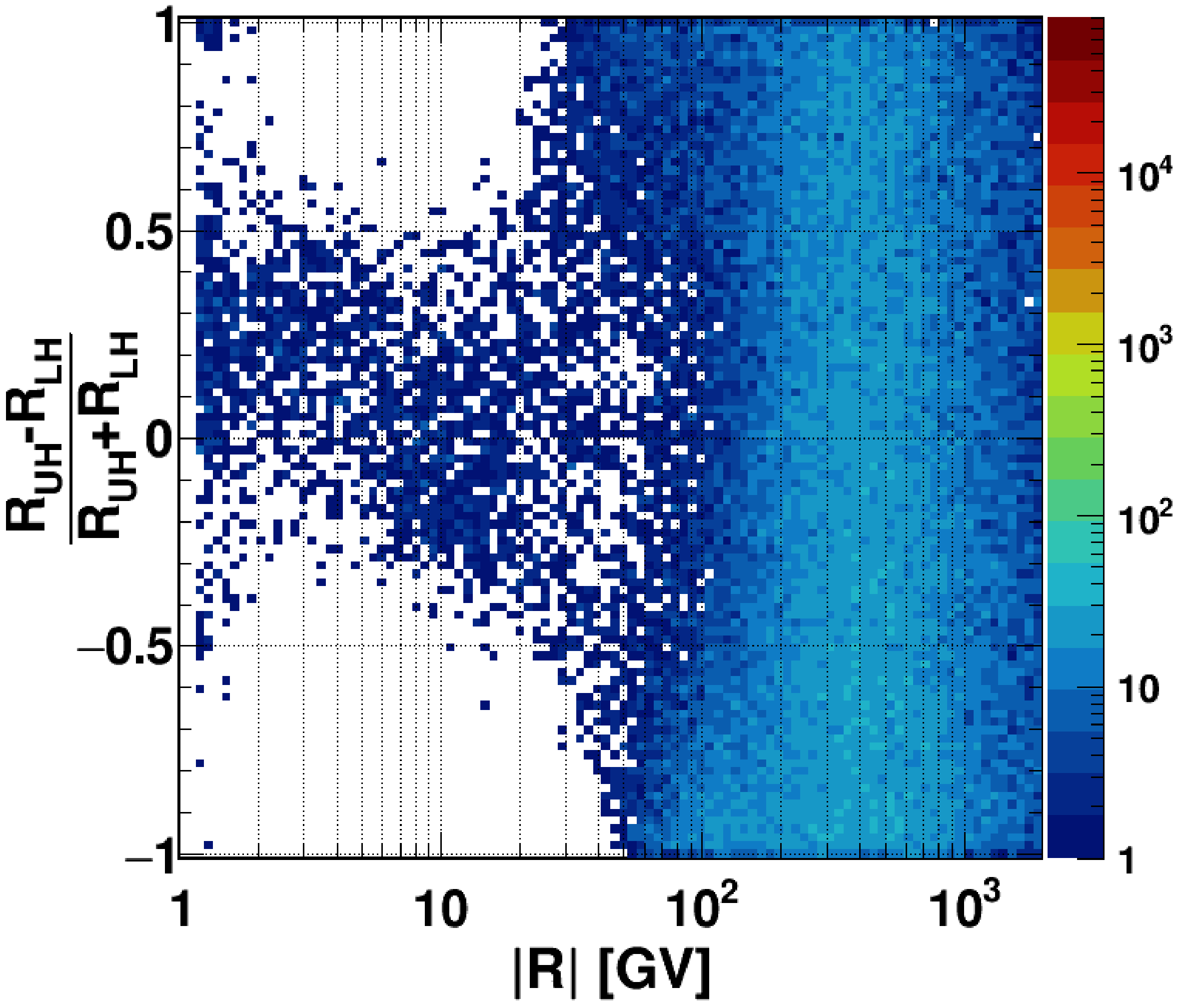


MC He4 $R_{INNER} > 0$

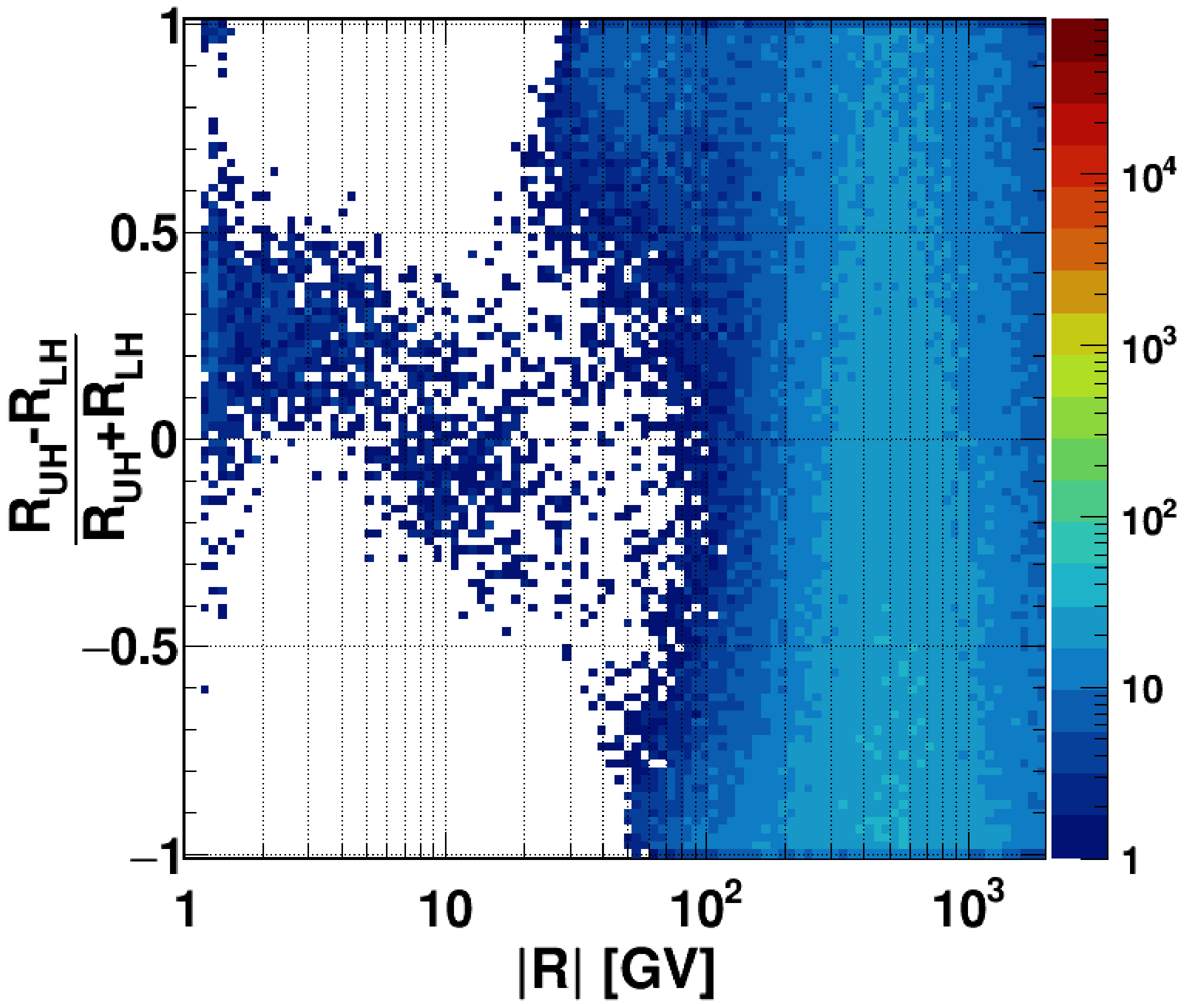


Classifier label dependency from Rigidity (R<0)

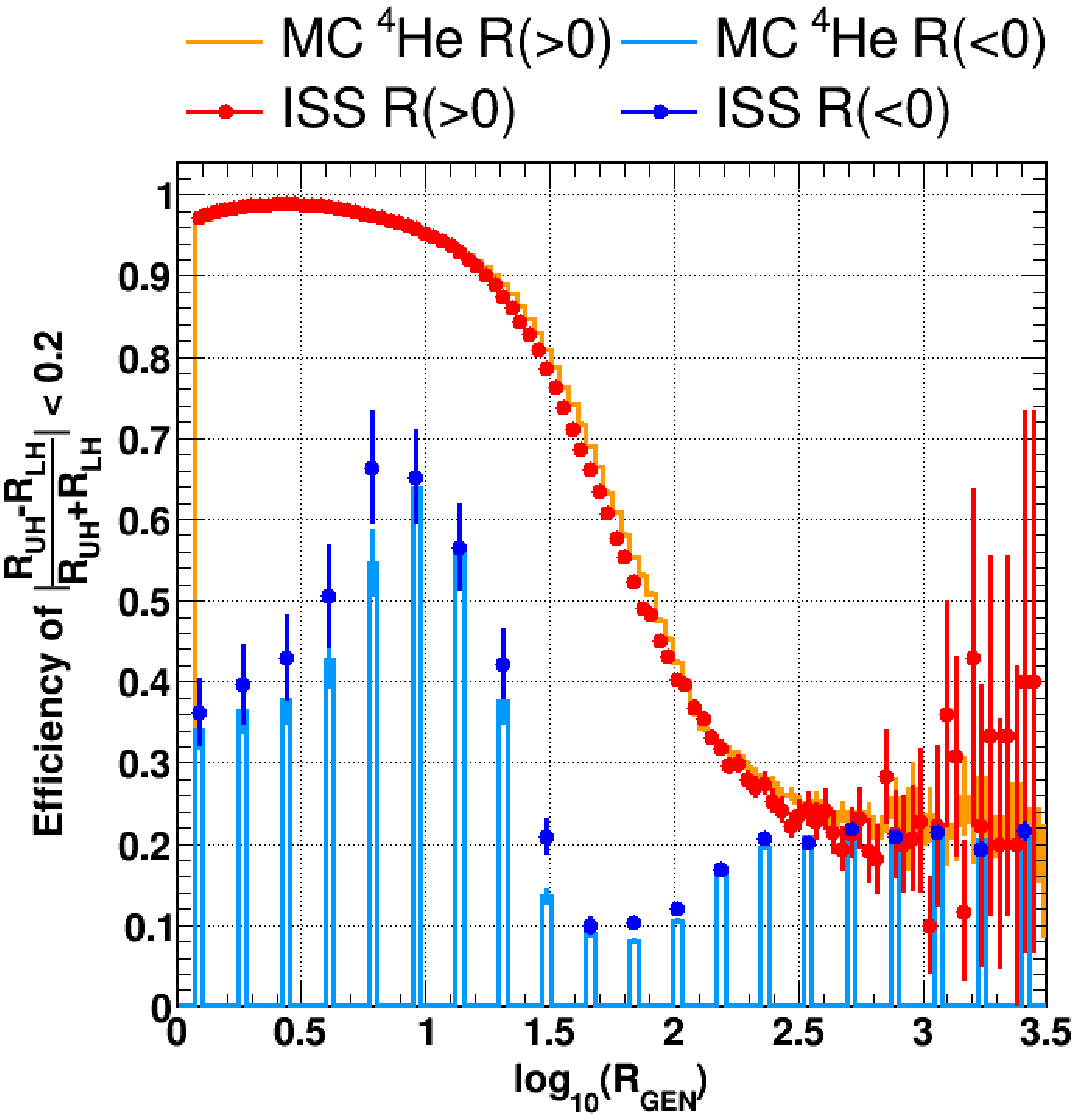
ISS data $R_{INNER} < 0$



MC He4 $R_{INNER} < 0$



Signal efficiency as a function of measured rigidity



Supervised classifier input features and training

Input features:

- **Identify variables with high classification power**
- 23 input variables

Inner Tracker:

- TRK Y residual $(\times\ 6)^*$
- TRK X residual $(\times\ 3)^{**}$
- TRK Y hit number
- TRK X hit number
- $\frac{|R_{INNER}|-|R_{INNERL1}|}{|R_{INNER}|+|R_{INNERL1}|}$
- MIN and MAX Z on TRK plane

* no L8
** L3/L5/L6

RICH:

- Z_{RICH}
- N. PMTs (ring)
- N. Photo-electrons

Training on R>0:

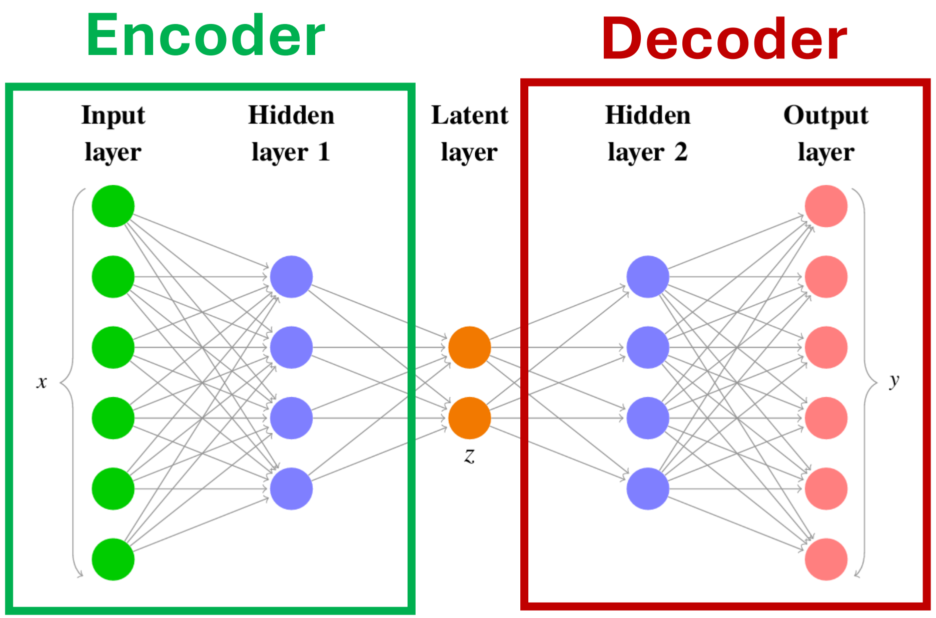
- Sample balancing: "signal events" are shuffled and downsampled.
- Afterwards, $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$ fraction is 50 % and $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| > 0.2$ fraction is 50 %.
  - ISS data (2.8×10^6) and MC (4.9×10^6) events
  - 70% used for training, 30% for validation
- After the training CL_{ISS} is applied to R<0 ISS data, CL_{MC} is applied to R<0 MC.

Autoencoder

- Goal: **minimise losses in the compression-decompression process.**
  - Reconstruction error : $\mathrm{MSE} = \frac{(\boldsymbol{X}-\boldsymbol{Y})^2}{N_{features}}$

    $\boldsymbol{X}$ = input
    $\boldsymbol{Y}$ = AE's output

  - Well-reconstructed events dominate the R>0 sample.
  - Poorly reconstructed events dominate the R<0 sample.



- The autoencoder fails on badly reconstructed events, while well-reconstructed events are unaffected.

To have a variable defined between $[0.5,1]$, the sigmoid function ($\frac{1}{1+e^{-MSE}}$) is used:

$$\textbf{Anomaly score} = \frac{\mathbf{1}}{\mathbf{1} + e^{-MSE}}$$

Reliable $\overline{He}$ candidates should have an anomaly score ~ 0.5.

Autoencoder input features and training

Input features:

- **Identify variables with high classification power between the R>0 and the R<0 sample**
- 30 input variables

Training on R>0:

- Cut on reconstructed rigidity: $R_{INNER} < 200$ [GV] (MDR).
  - ISS data (1.2×10^7) and MC (1.2×10^7) events
  - 70% used for training, 30% for validation

After the training AE_{ISS} is applied to R<0 ISS data, AE_{MC} is applied to R<0 MC.

Inner Tracker:

- TRK Y residual $(\times 7)$
- TRK X residual $(\times 4)^*$
- TRK Y hit number
- TRK X hit number
- $\frac{|R_{INNER}| - |R_{INNERL1}|}{|R_{INNER}| + |R_{INNERL1}|}$
- $\frac{R_{UH} - R_{LH}}{R_{UH} + R_{LH}}$
- TRK NormEdep2XY $(\times 4)^{**}$
- MIN – MAX CH on TRK plane

ACC:

- ACC counters

Time Of Flight:

- L3 on-time cluster
- L4 on-time cluster

* L3/L5/L7/L8

** L5/L6/L7/L8

$$\text{NormEdep2Y} = \frac{E_{on-track}\,Y}{E_{dep}\,Y(2\,cm) + E_{on-track}Y}$$

$$^{**}\text{NormEdep2XY} = \frac{(\text{NormEdep2Y} - \text{NormEdep2X})}{(\text{NormEdep2Y} + \text{NormEdep2X})}$$

Workflow:

- The networks have information regarding the rigidity, but not β (TOF, NaF, AGL).
- After the training (on ISS and MC), divide into 11 rigidity bins:

[1.92,4.02], [4.02,6.47],[6.47,9.26],[9.26,13],[13,18],

[18,24.7],[24.7,31.1],[31.1,36.1],[36.1,41.9],[41.9,48.5], 48.5<

- For each rigidity bin, cut on the network's outputs using the efficiency on "signal-like events".
- Reconstructs the mass of the survivors (ISS and MC).

Using R>0 sample (Data and MC), for each detector (TOF, NaF, AGL) and rigidity bin get:

- CL score for $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$
- AE anomaly score for $R_{INNER} > 0.$
- Cut values at predefined efficiencies.

Apply the cuts on R<0 sample (Data and MC).

Cutting on the anomaly score changes the distribution of the classifier score?

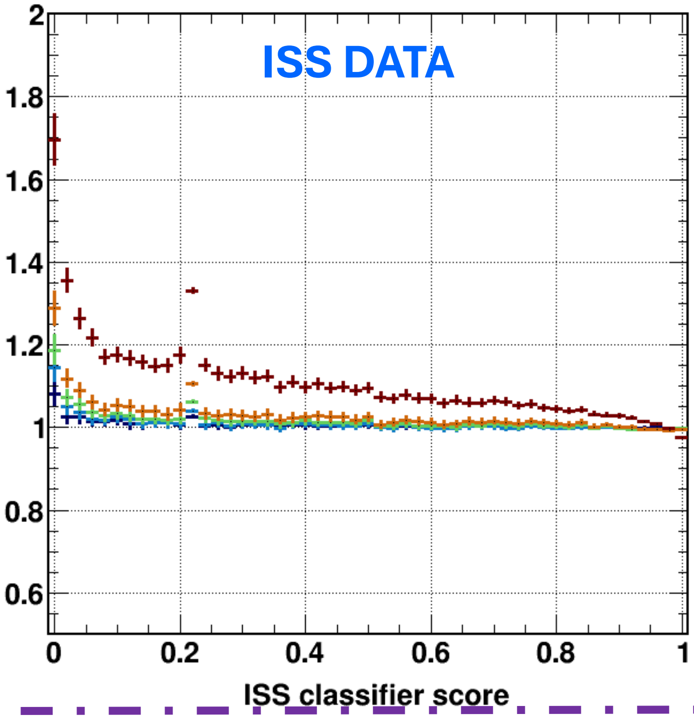
Are the outputs of the network independent and orthogonal?

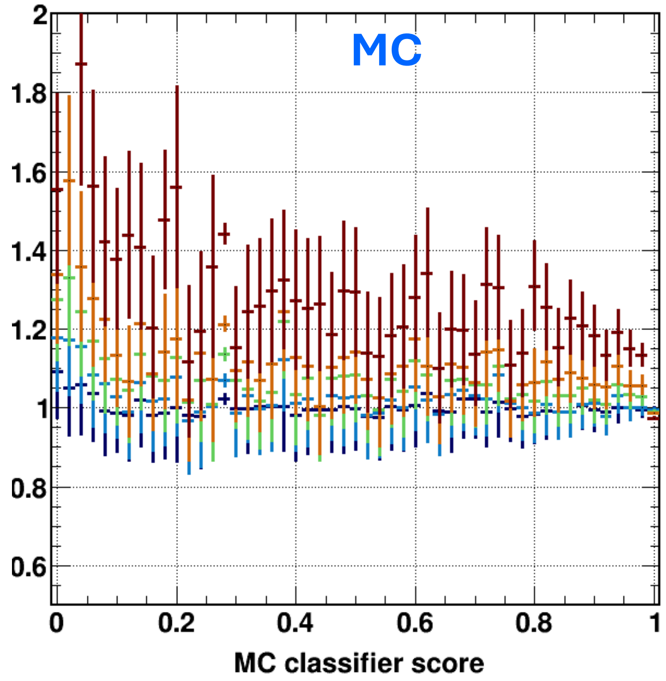
Check on the orthogonality TOF $R \in [1.92, 4.02]$ GV

Ratio: no cut / cut on score

eff. 0.95 eff. 0.90 eff. 0.85
eff. 0.80 eff. 0.67

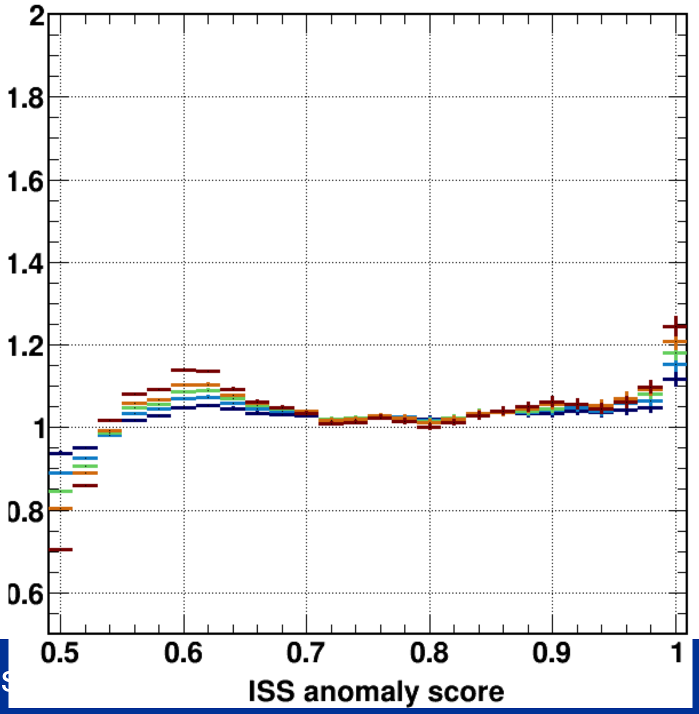
CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

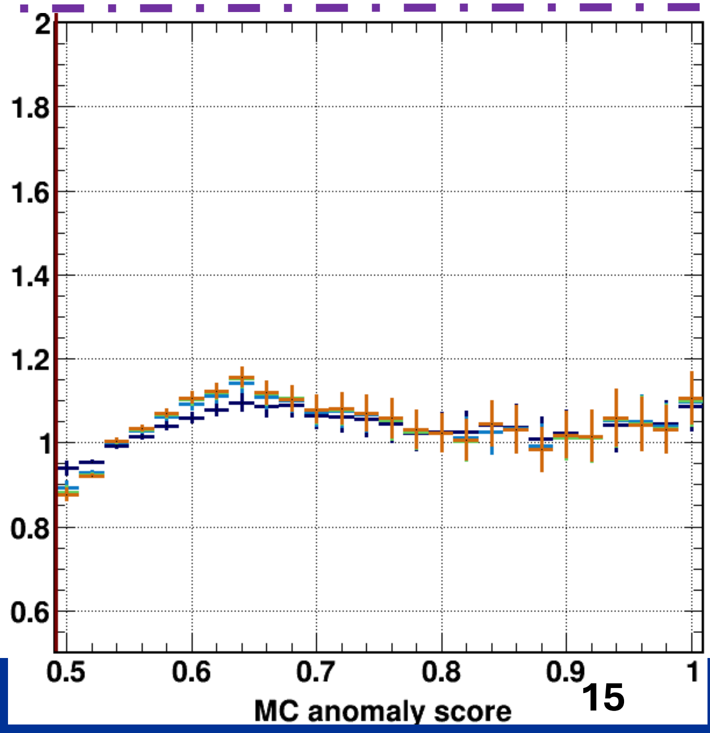




AE anomaly score $R_{INNER} > 0$

eff. 0.95 eff. 0.90 eff. 0.85
eff. 0.80 eff. 0.67



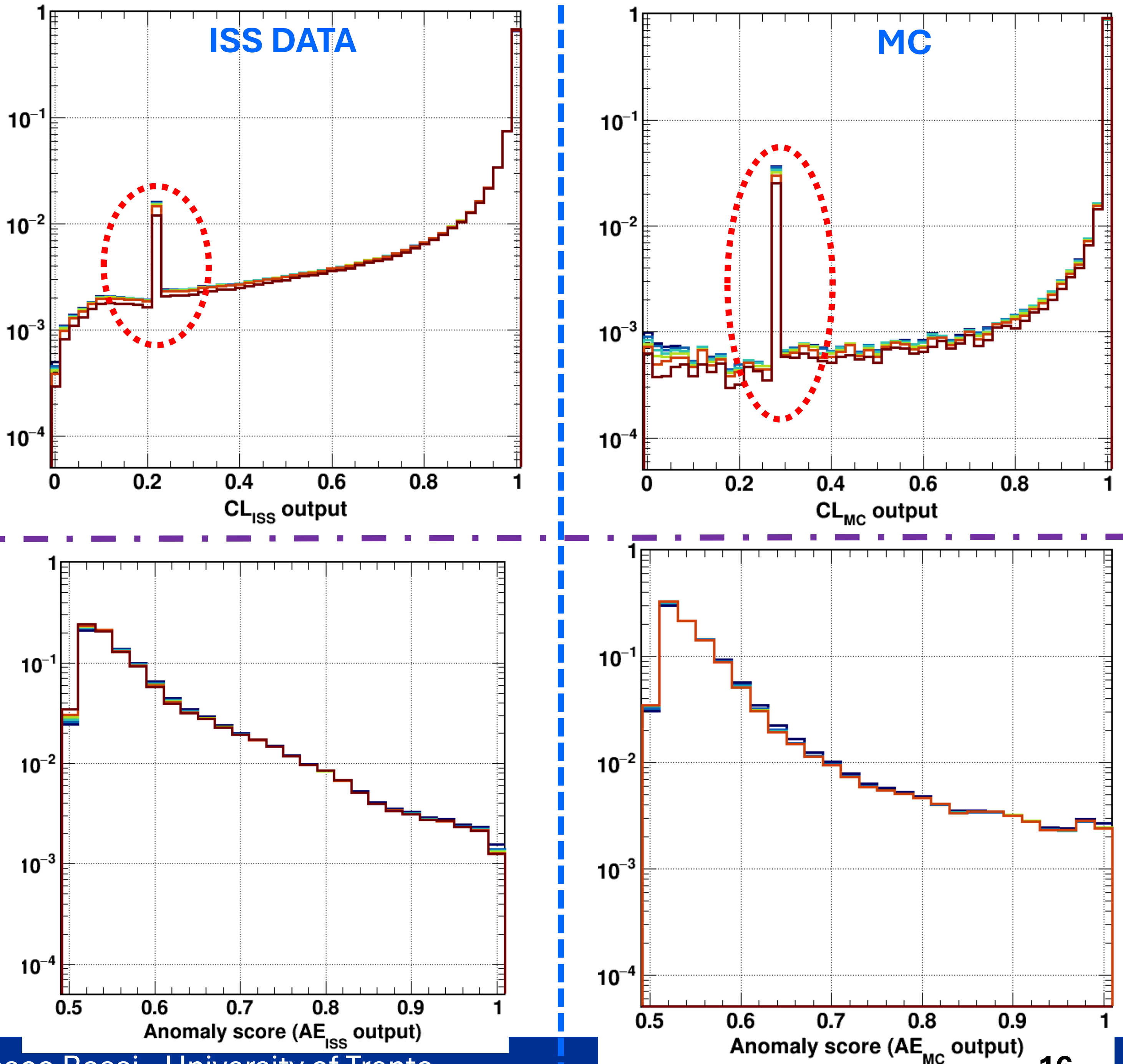


Check on the orthogonality TOF $R \in [1.92, 4.02]$ GV

eff. 0.95 eff. 0.90 eff. 0.85
eff. 0.80 eff. 0.67

CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

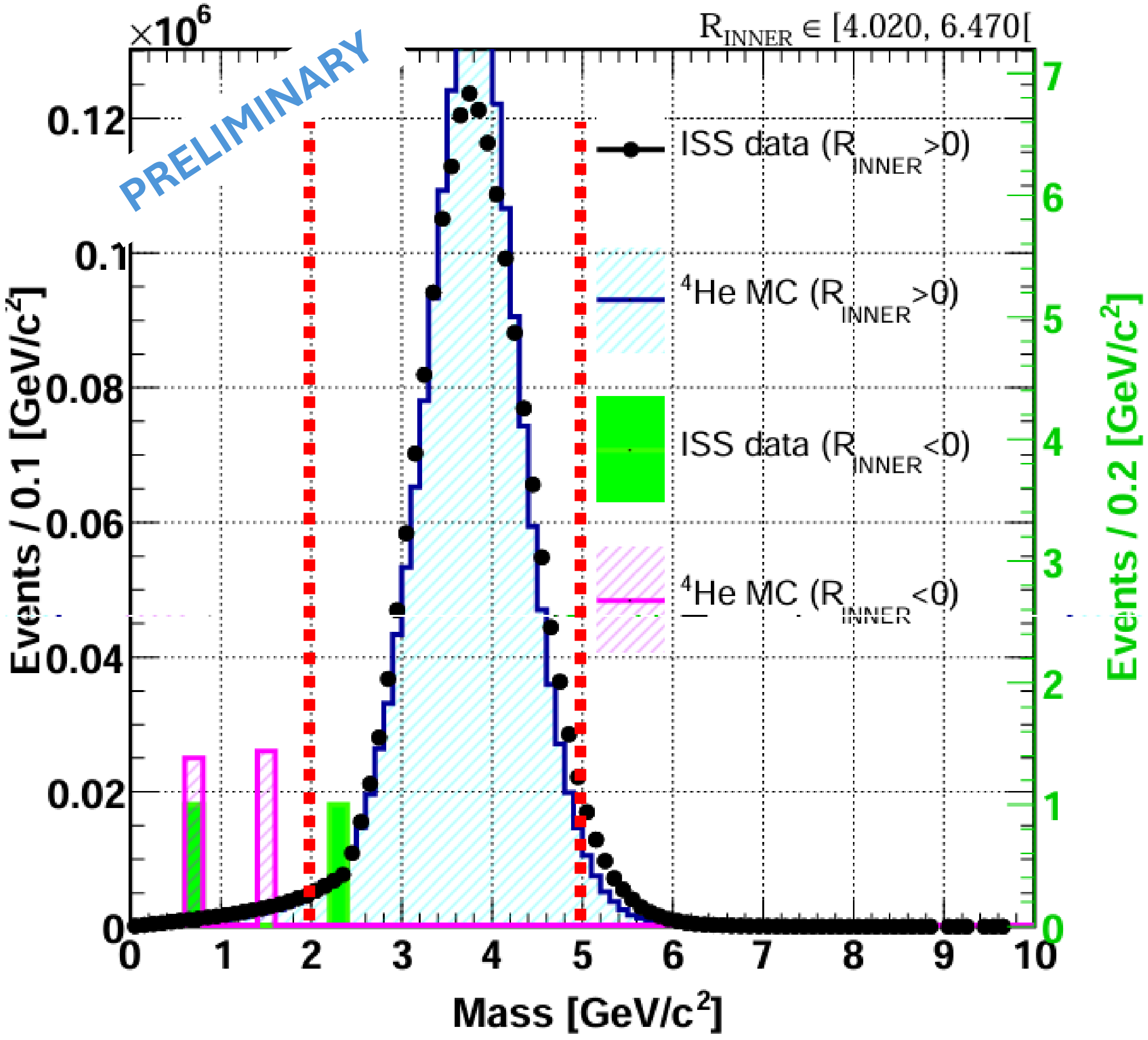
AE anomaly score $R_{INNER} > 0$



The networks are coherent.
Cutting on the scores of one network increases the % of “good events” in the score distribution of the other.

Mass distribution TOF
$R \in [1.92, 4.02]$ GV

- 80% efficiency on R>0 AE anomaly score
- 75% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

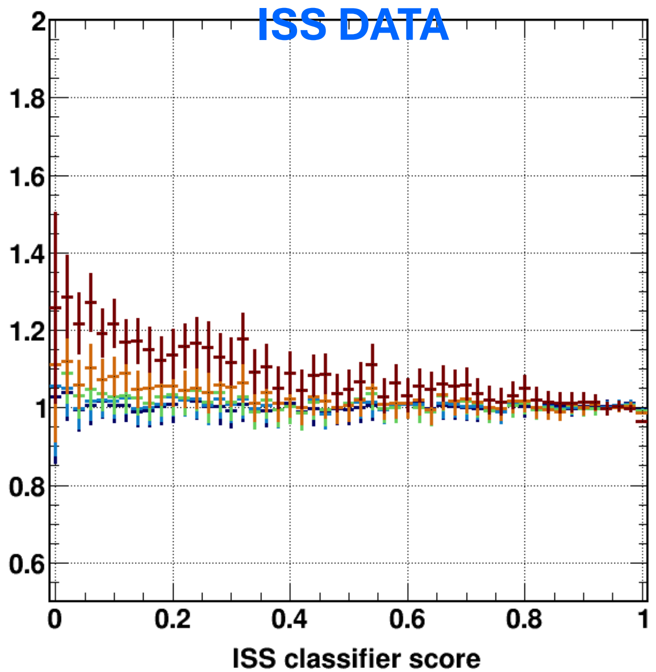


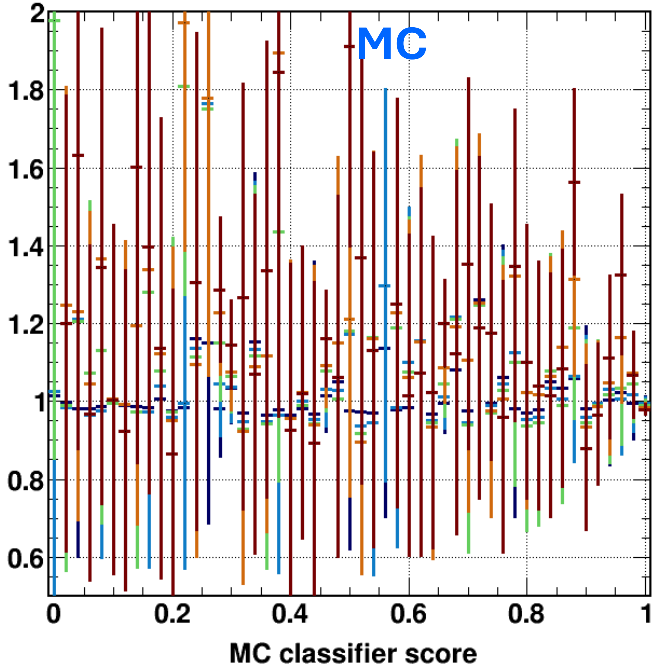
Check on the orthogonality NaF $R \in [6.47, 9.26]$ GV

Ratio: no cut / cut on score

eff. 0.95 — eff. 0.90 — eff. 0.85
eff. 0.80 — eff. 0.67

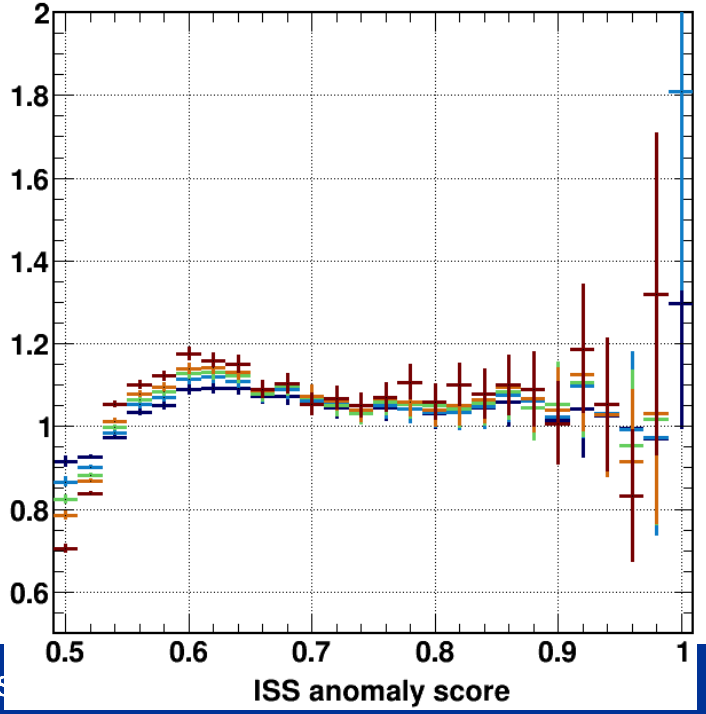
Classifier score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

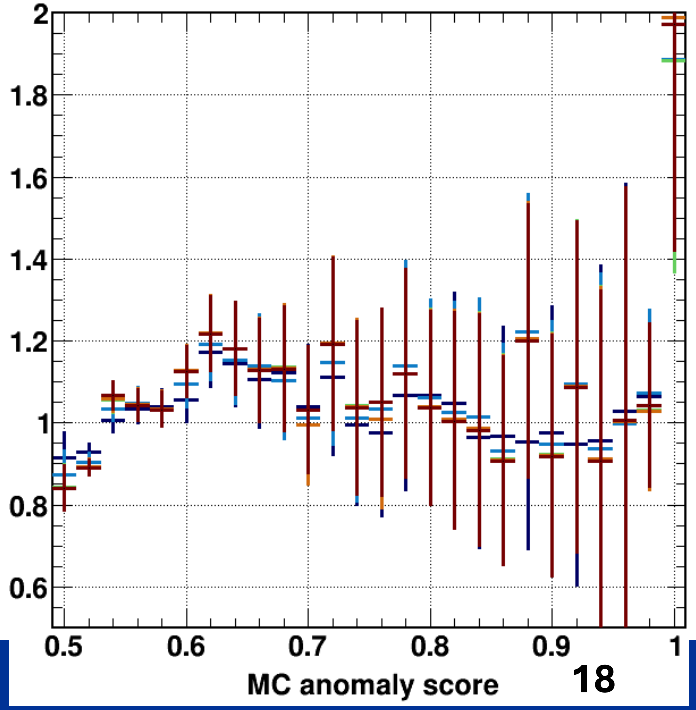




Anomaly score $\mathrm{R_{INNER}} > 0$

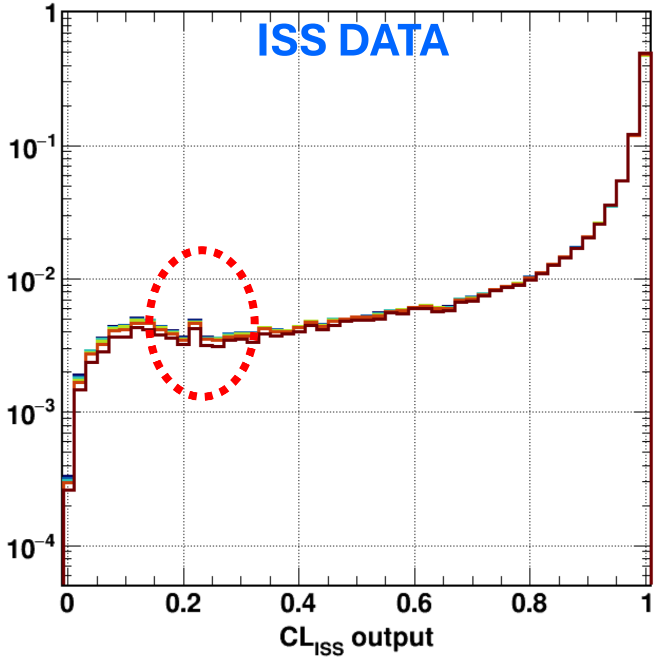
eff. 0.95 — eff. 0.90 — eff. 0.85
eff. 0.80 — eff. 0.67

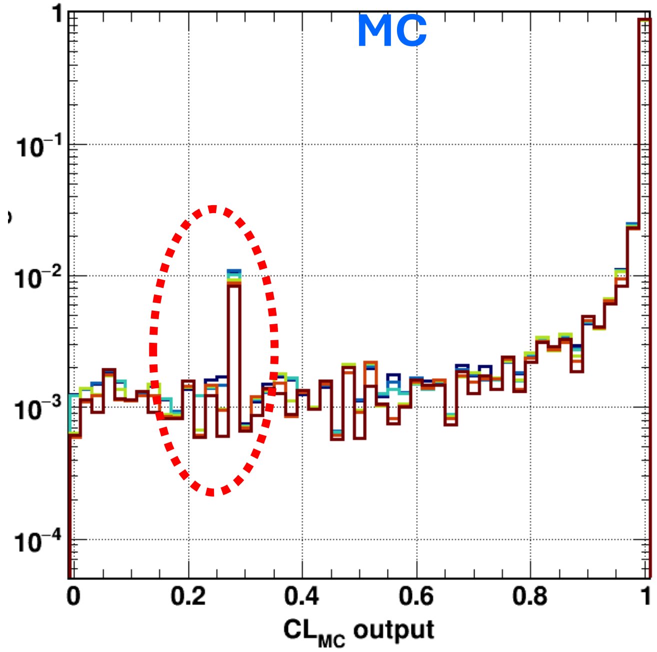




Check on the orthogonality NaF $R \in [6.47, 9.26]$ GV

— Good reco. (R_{INNER}>0) — eff. 0.95 — eff. 0.90
— eff. 0.85 — eff. 0.80 — eff. 0.67

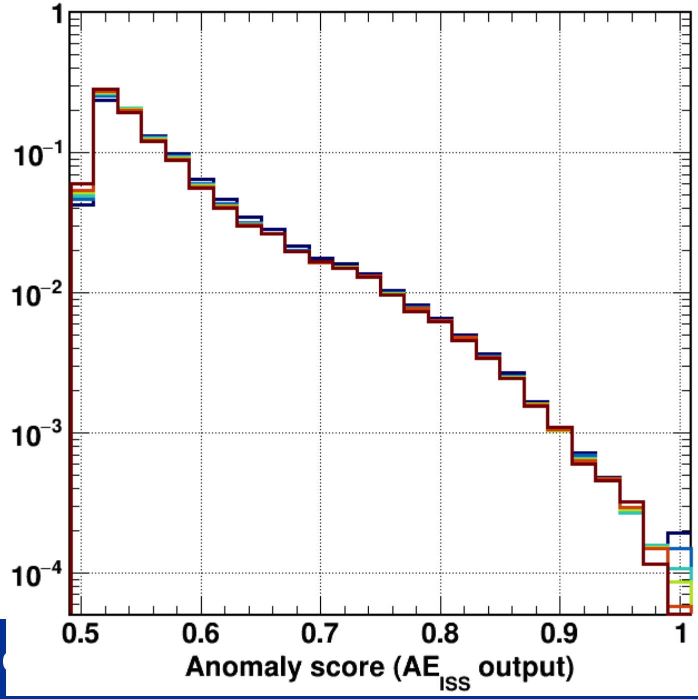


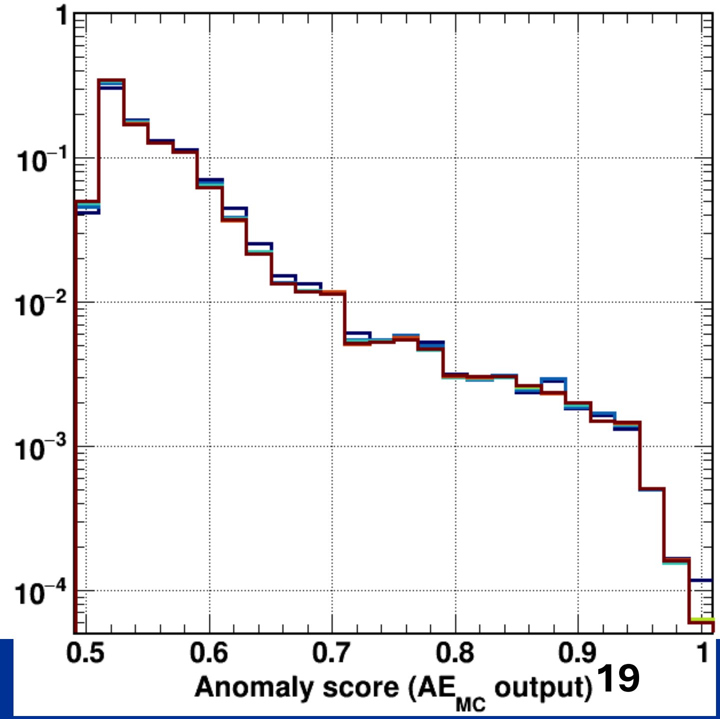


Classifier score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

Anomaly score $\mathrm{R_{INNER}} > 0$

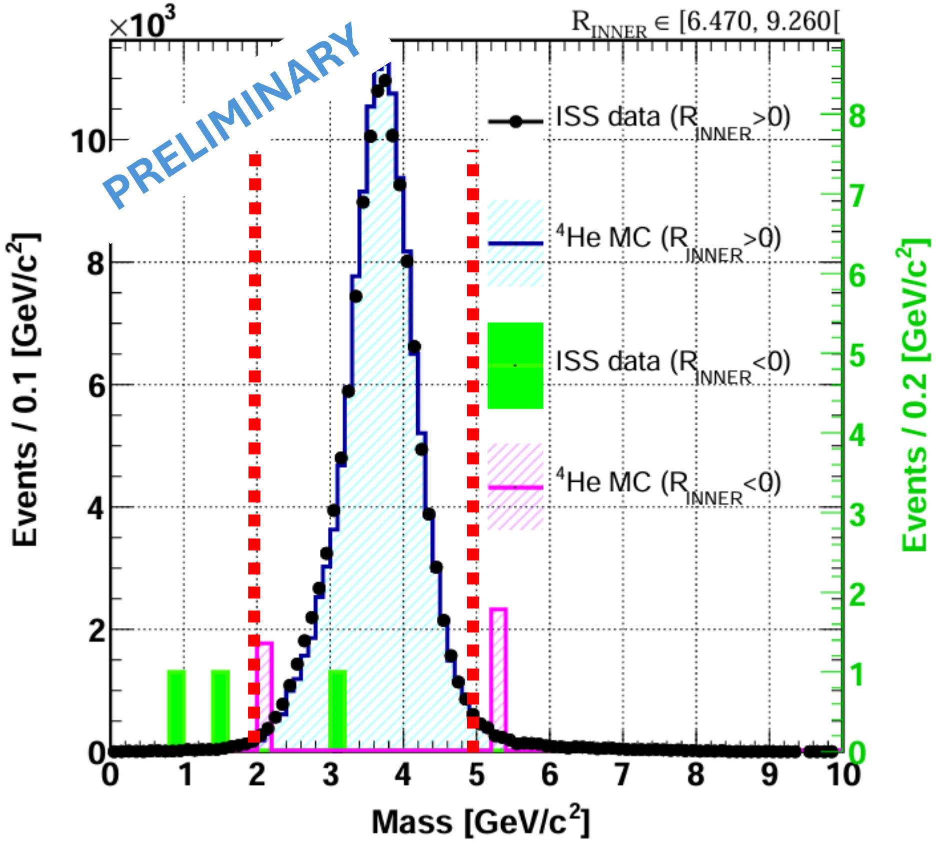
— ISS data (R_{INNER}>0) — eff. 0.95 — eff. 0.90
— eff. 0.85 — eff. 0.80 — eff. 0.67



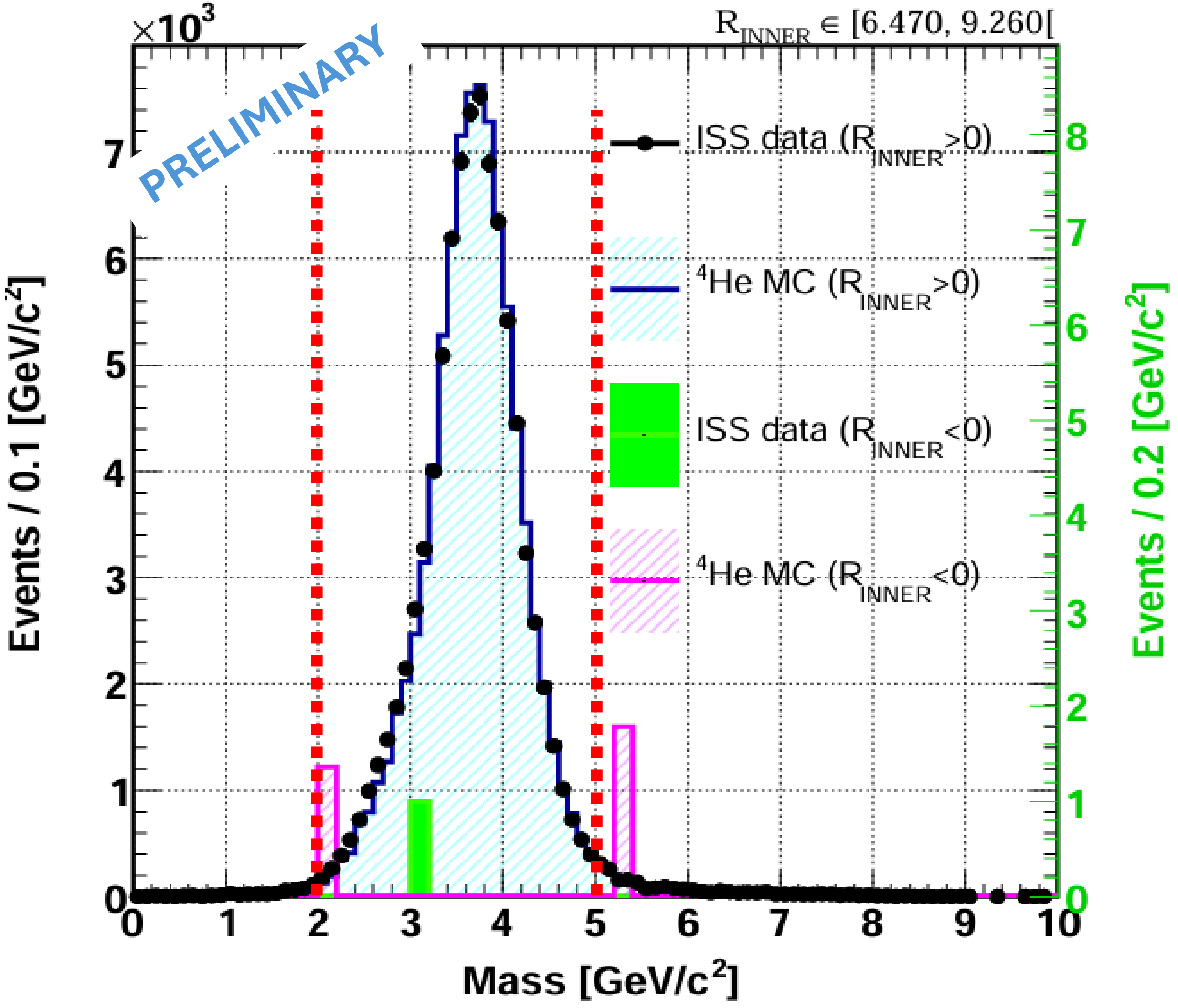


Mass distribution NaF
$R \in [6.47, 9.26]$ GV

- 80% efficiency on R>0 AE anomaly score
- 75% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$



- 80% efficiency on R>0 AE anomaly score
- 50% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

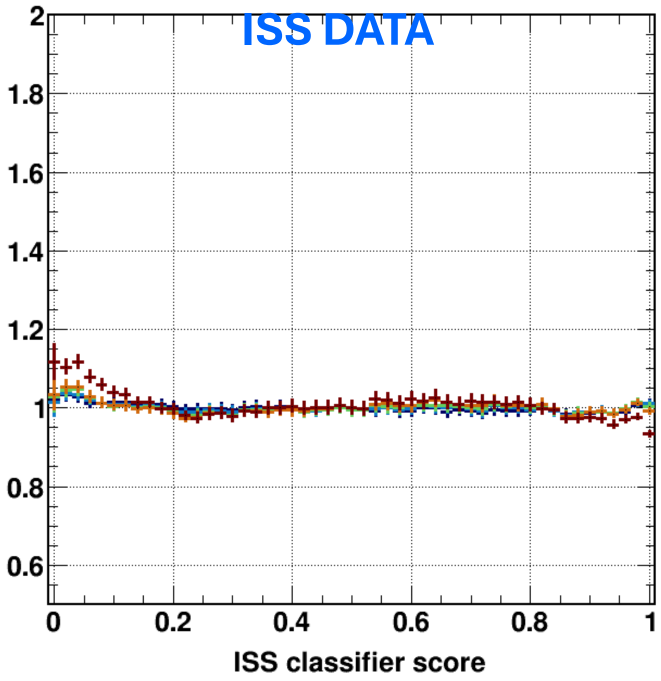


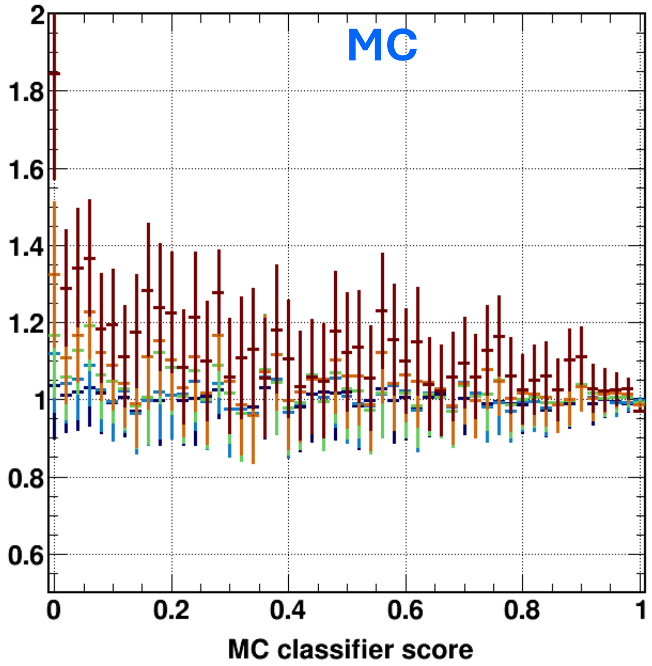
Check on the orthogonality AGL $R \in [18, 24.7]$ GV

Ratio: no cut / cut on score

eff. 0.95 eff. 0.90 eff. 0.85
eff. 0.80 eff. 0.67

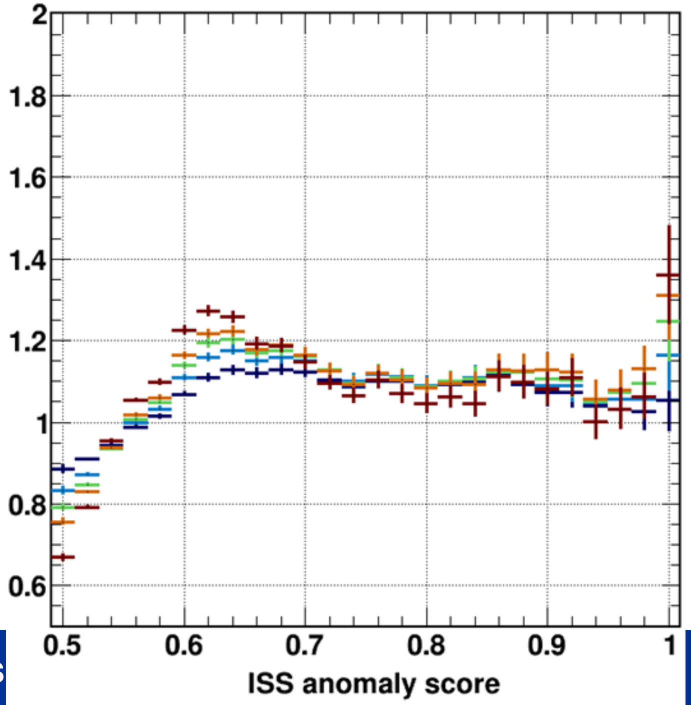
Classifier score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

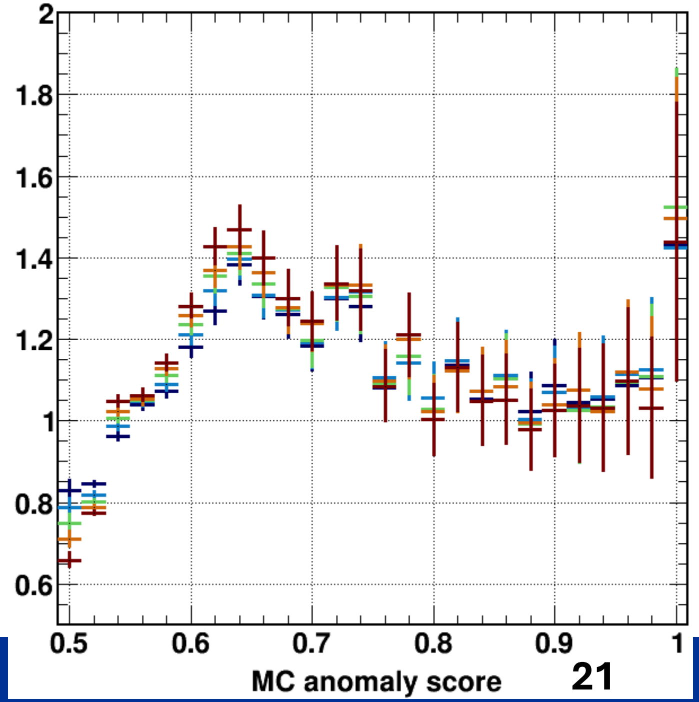




Anomaly score $\mathrm{R_{INNER}} > 0$

eff. 0.95 eff. 0.90 eff. 0.85
eff. 0.80 eff. 0.67

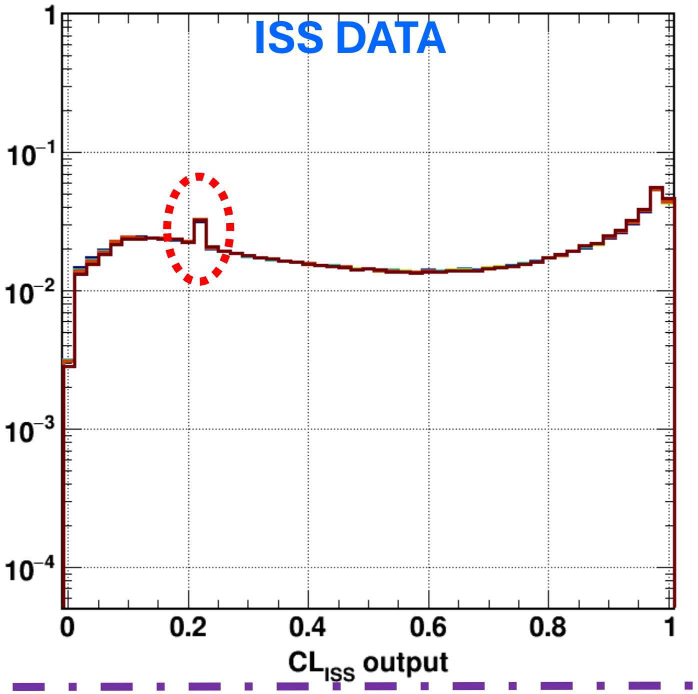


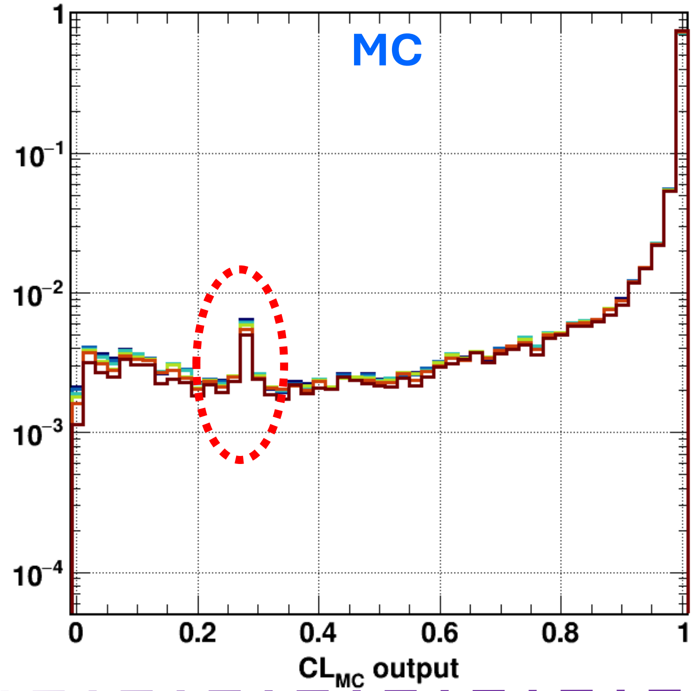


Check on the orthogonality AGL $R \in [18, 24.7]$ GV

— Good reco. (R_{INNER}>0) — eff. 0.95 — eff. 0.90
— eff. 0.85 — eff. 0.80 — eff. 0.67

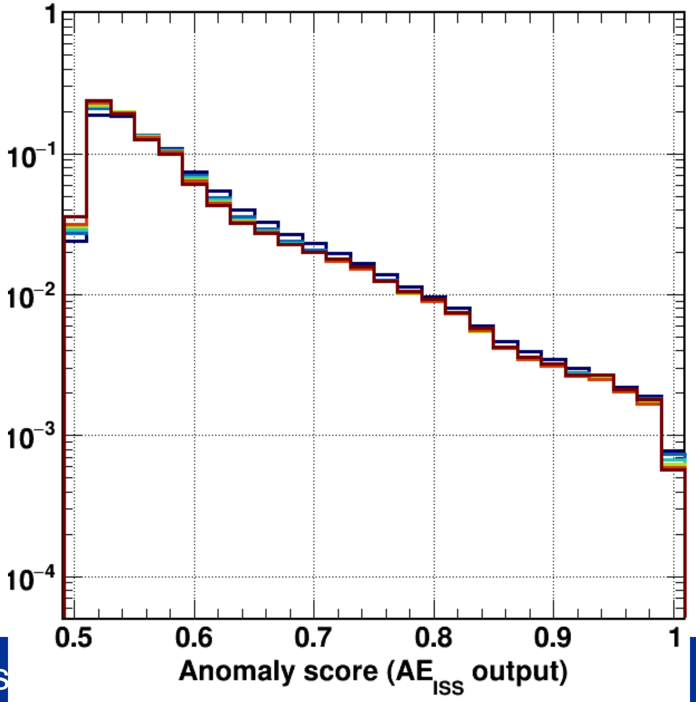
Classifier score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

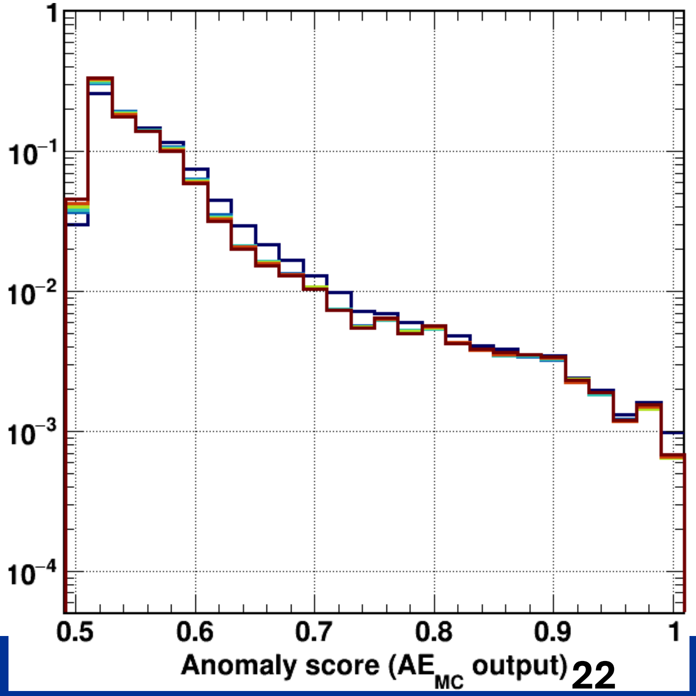




Anomaly score $R_{INNER} > 0$

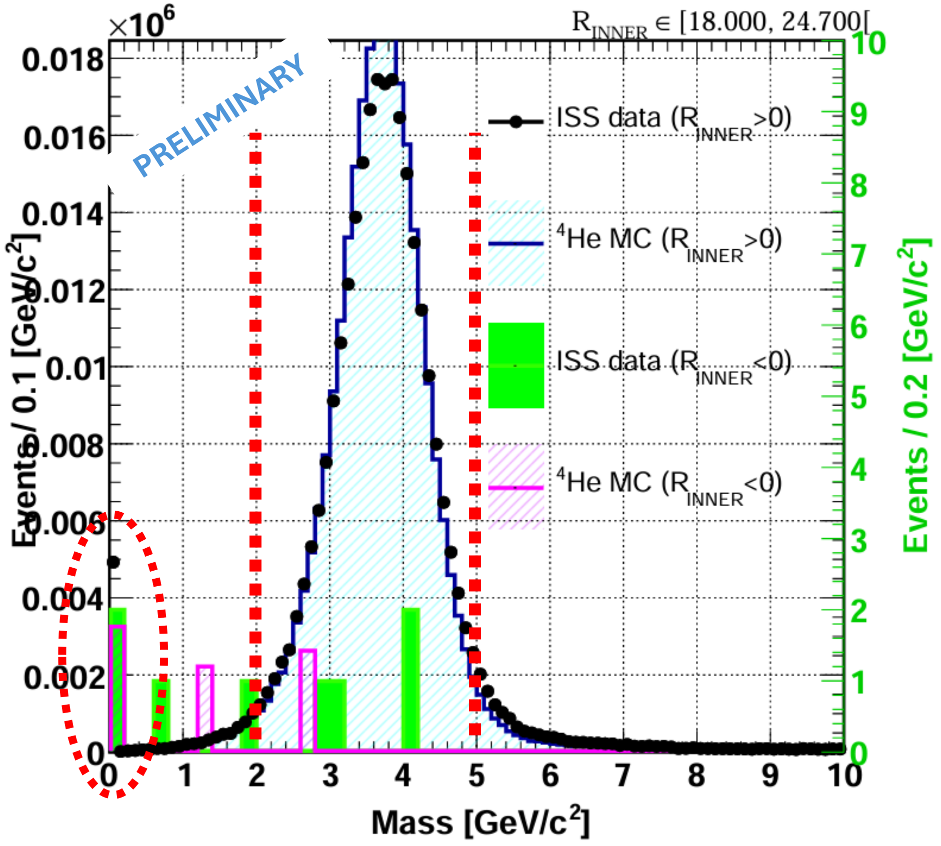
— ISS data (R_{INNER}>0) — eff. 0.95 — eff. 0.90
— eff. 0.85 — eff. 0.80 — eff. 0.67



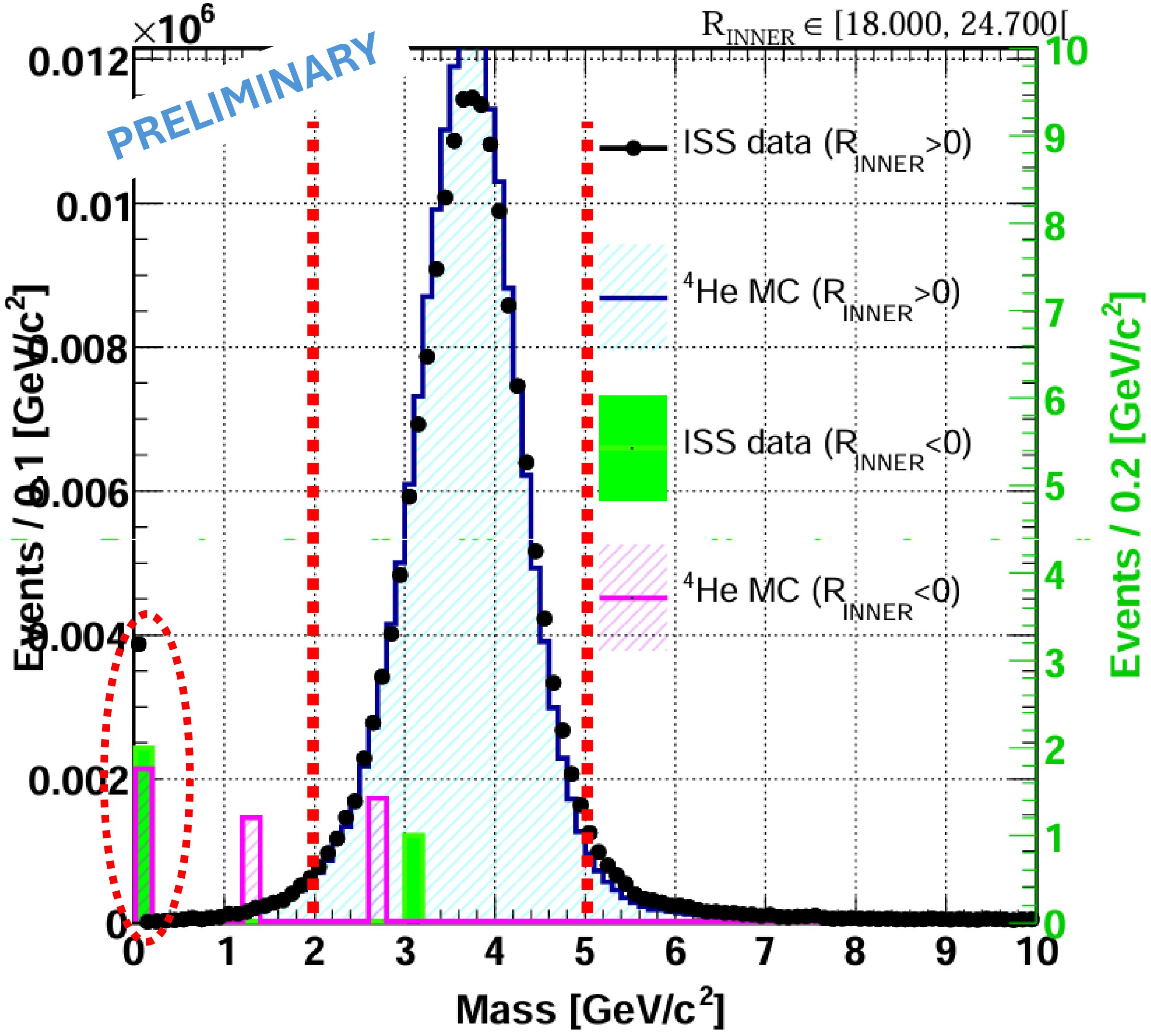


Mass distribution AGL
$R \in [18, 24.7]$ GV

- 80% efficiency on R>0 AE anomaly score
- 75% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$



- 80% efficiency on R>0 AE anomaly score
- 50% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$

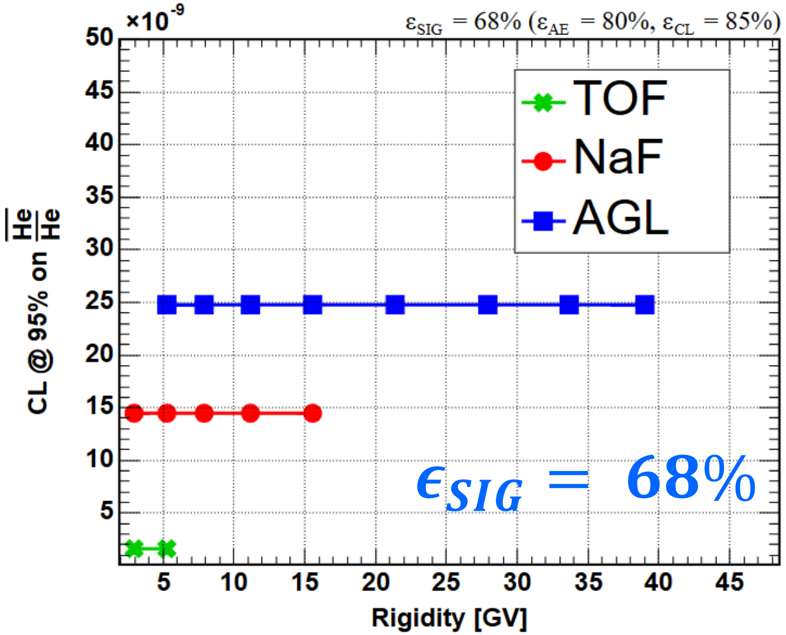


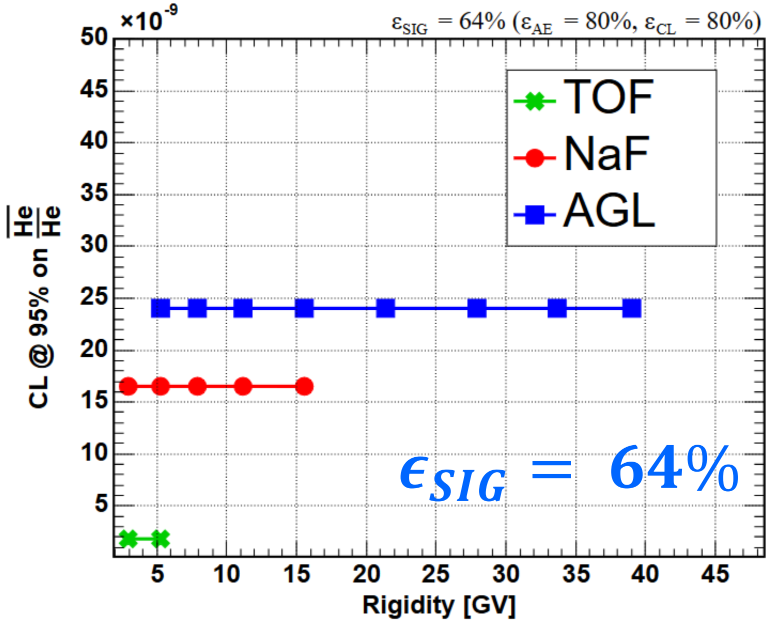
Confidence limit estimate

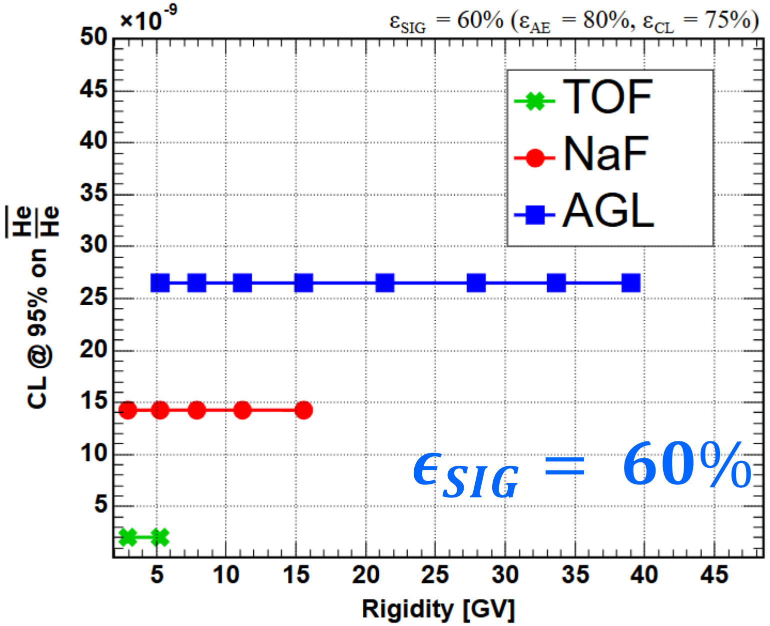
- Identify $\overline{He}$ candidates:
  - additional cut on the reconstructed mass m_{REC} ∈ [2,5] GeV/c^2
- Number of survivors in data (n_{ISS}) and MC ($\mathrm{n_{MC}}$), integrated in R.
- Monte Carlo toy procedure.
- Repeat it for
  - TOF:
    - [1.92,4.02], [4.02,6.4]
  - NaF:
    - [1.92,4.02], [4.02,6.4], [6.47,9.26],[9.26,13],[13,18]
  - AGL:
    - [4.02,6.4],[6.47,9.26],[9.26,13],[13,18],[18,24.7], [24.7,31.1],[31.1,36.1],[36.1,41.9],[41.9,48.5], 48.5<

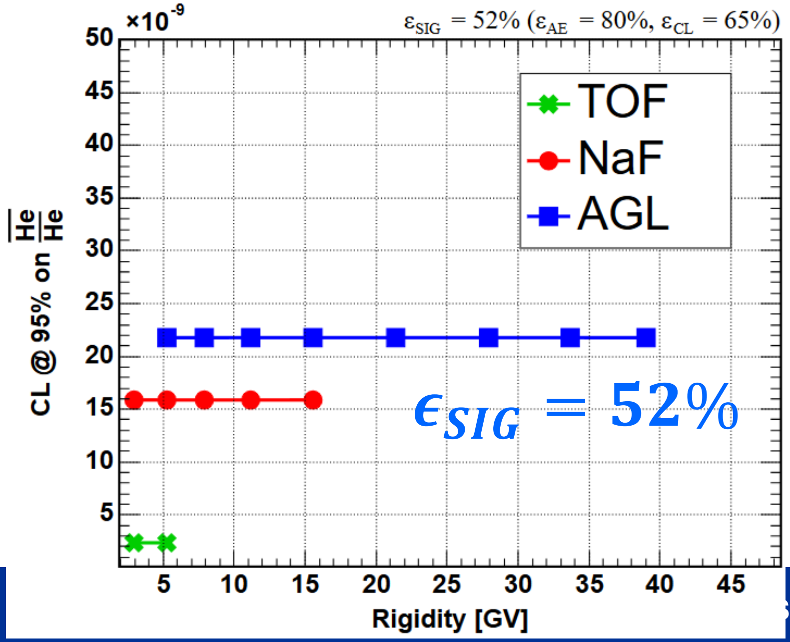
95% confidence limit on $\frac{\overline{He}}{He}$ at 80% eff. on AE anomaly score

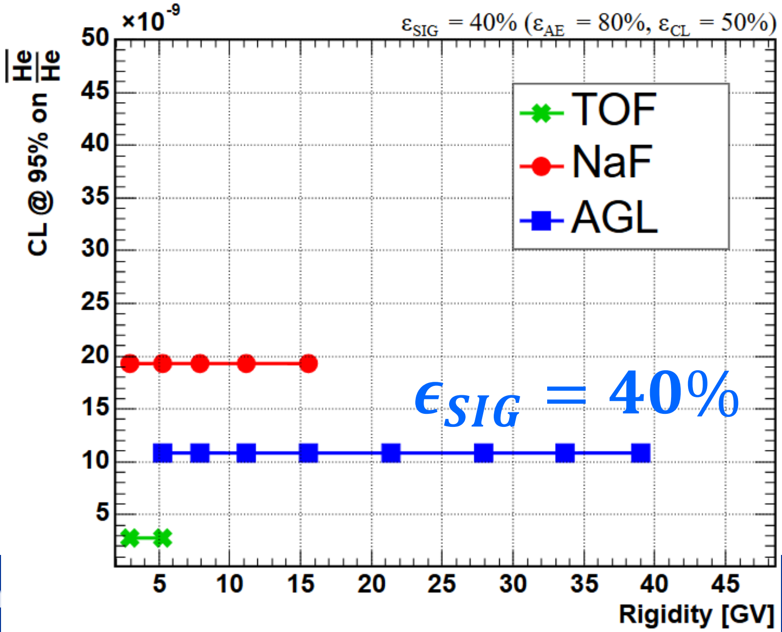
PRELIMINARY











Conclusions

- Two **independent and orthogonal** Neural Networks have been presented:
  1. Supervised classifier quantifies the quality of the reconstructed rigidity.
  2. Unsupervised autoencoder identifies well-reconstructed events.

- The networks trained on MC are used to study their performance and improve them.
- The networks trained on ISS are used to identify possible $\overline{He}$ candidates.
- **95% confidence limit on $\frac{\overline{He}}{He}$ for TOF, NaF and AGL detector.**

- Studies on the reliability of the $\overline{He}$ candidates (distance from tracker feet). **ONGOING**
- Cross-check with MIT candidates **ONGOING**
- Coverage of the MC toys procedure **ONGOING**
- Combination of the limits **ONGOING**

Event selection:

RTI

Good RTI
Outside SAA
Livetime > 0.5
Zenith $< 40\,°$

TRIGGER

Physical trigger

TOF β

$N.\beta$ clusters $\geq\ 4$
$\beta \quad > 0.3$
$\chi^2_{TIME} < 10$

Tracker fiducial volume cut:
L1: |R|<62cm, |Y|<47cm
L2: |R|<62cm, |Y|<40cm
L3: |R|<46cm, |Y|<44cm
L4: |R|<46cm, |Y|<44cm
L5: |R|<46cm, |Y|<36cm
L6: |R|<46cm, |Y|<36cm
L7: |R|<46cm, |Y|<44cm
L8: |R|<46cm, |Y|<44cm

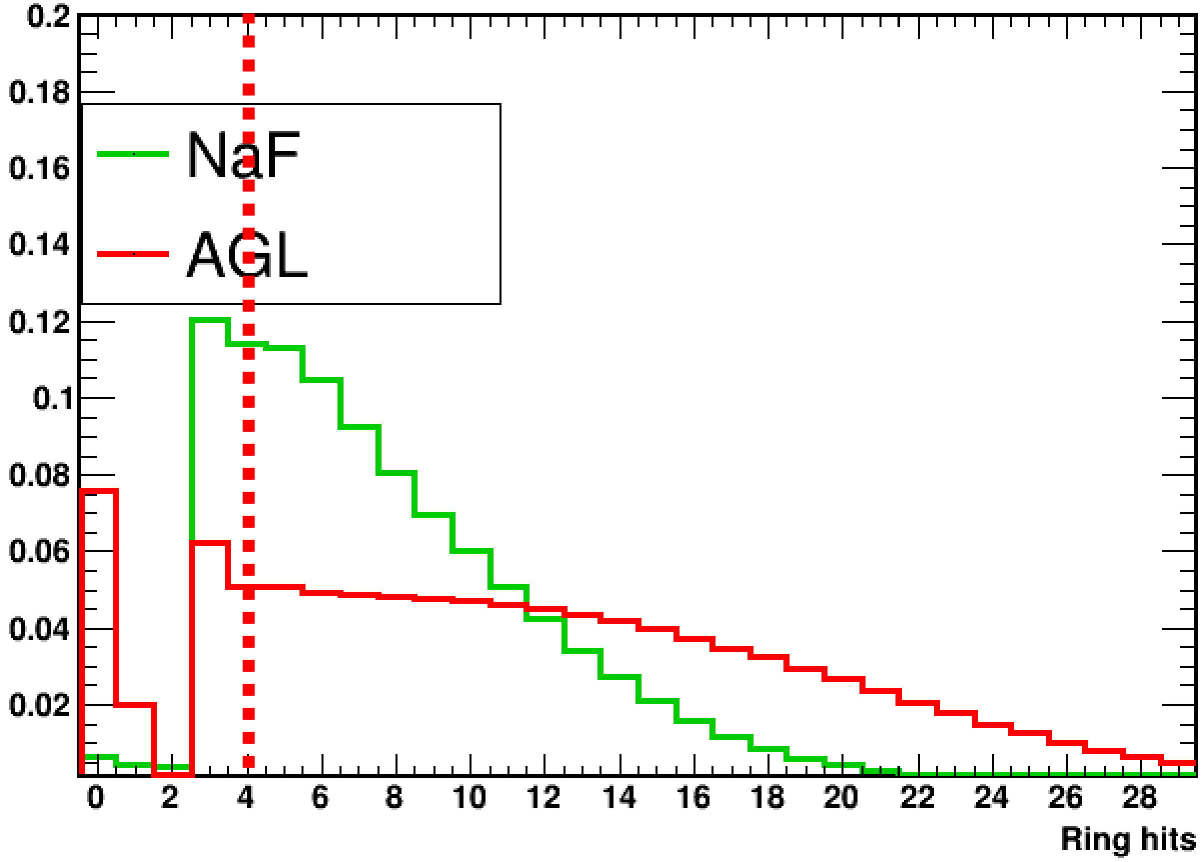
TRACK

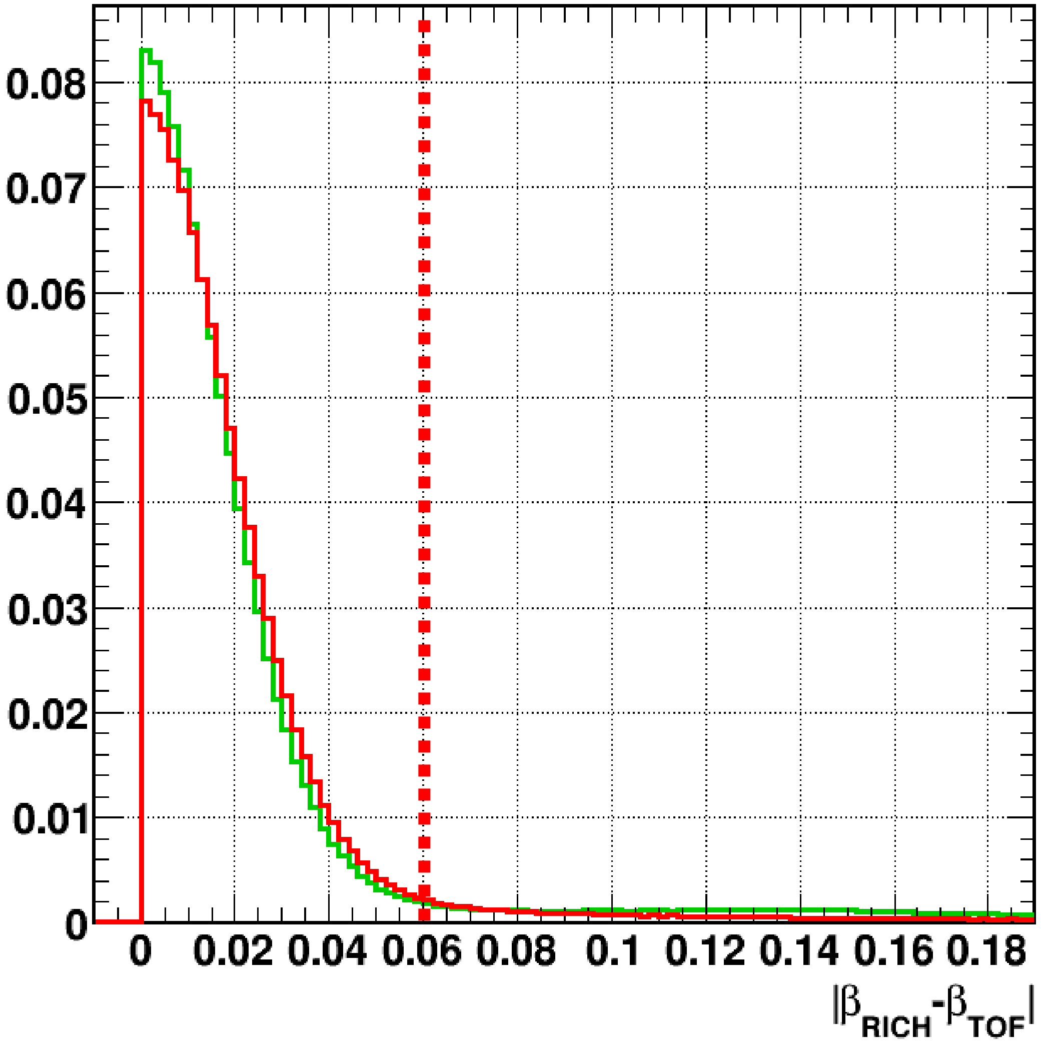
N. tracks ≤ 2
L2 XY hit
Y pattern: $\geq\ 5$ between L3-L8
X pattern: $\geq\ 4$, L3||L4, L5||L6, L7||L8

TRACKER

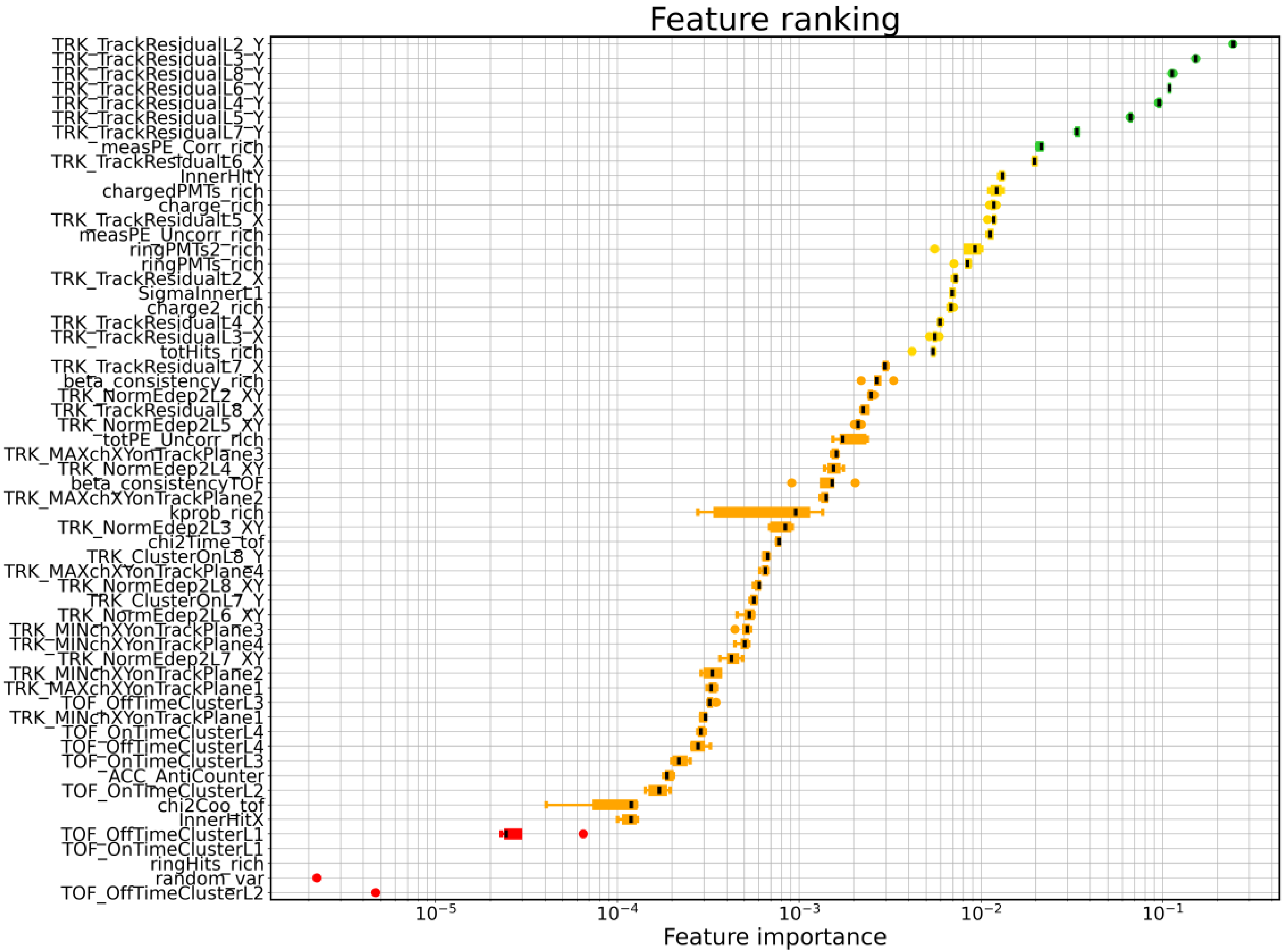
Z Inner Tracker $\in\ [1.7, 2.4]$
Inner fiducial volume
χ^2_Y (inner) < 10
χ^2_X (inner) < 10

Rich hits and compatibility with TOF

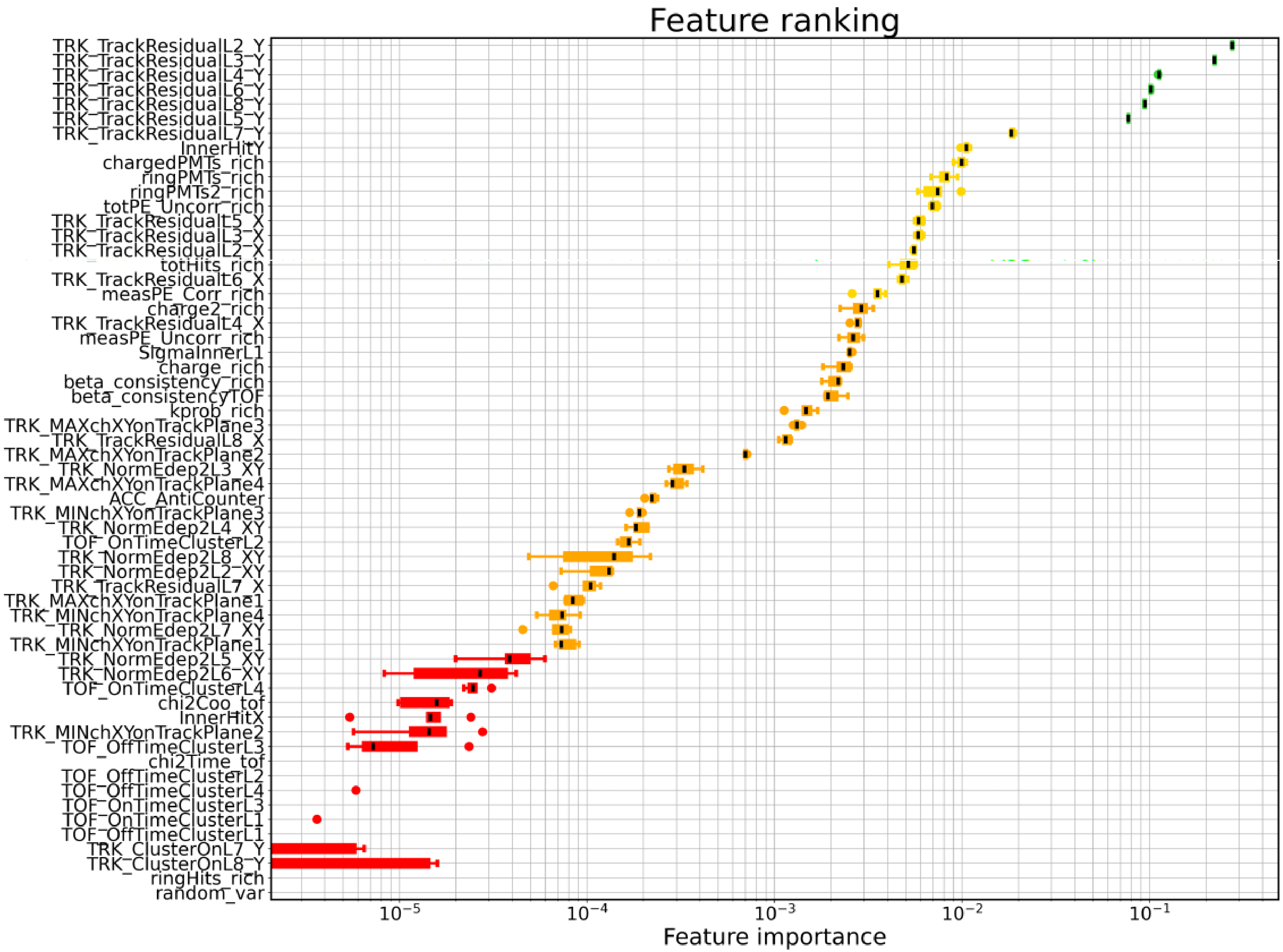




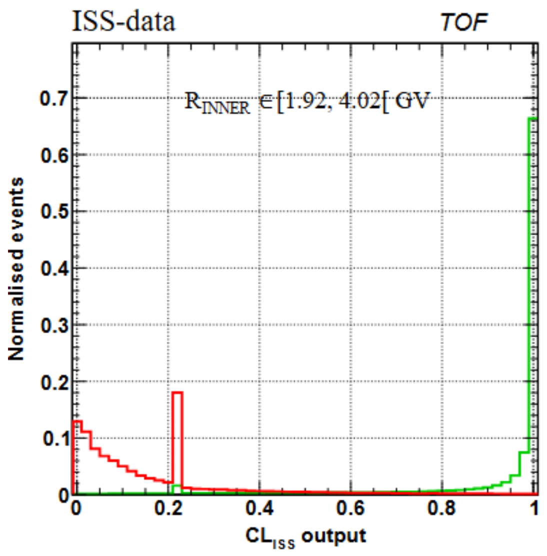
Feature Ranking CL_{ISS}



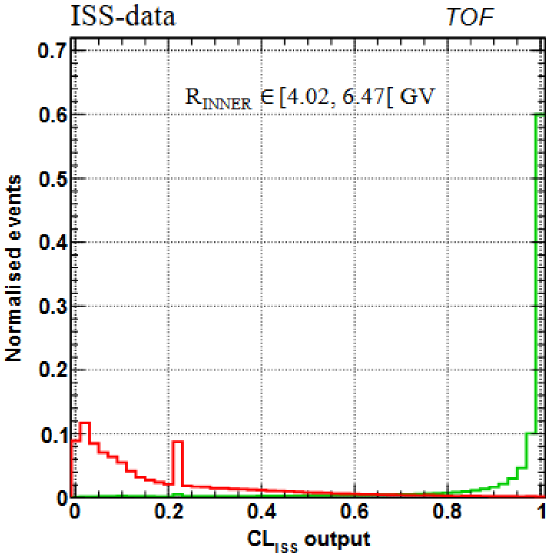
Feature Ranking CL_{MC}



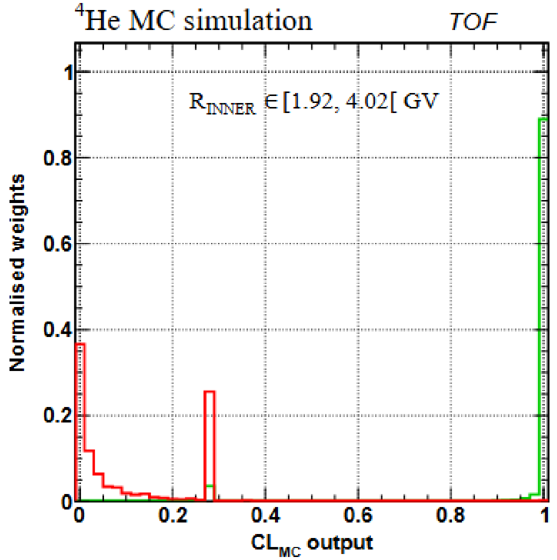
Scores TOF



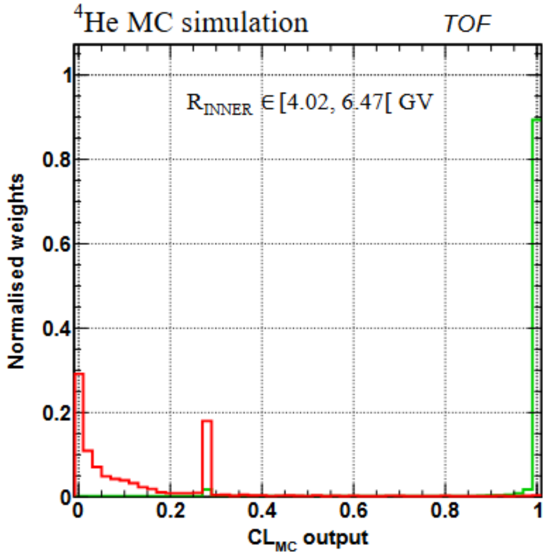
ISS-data
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised events
CL_{ISS} output



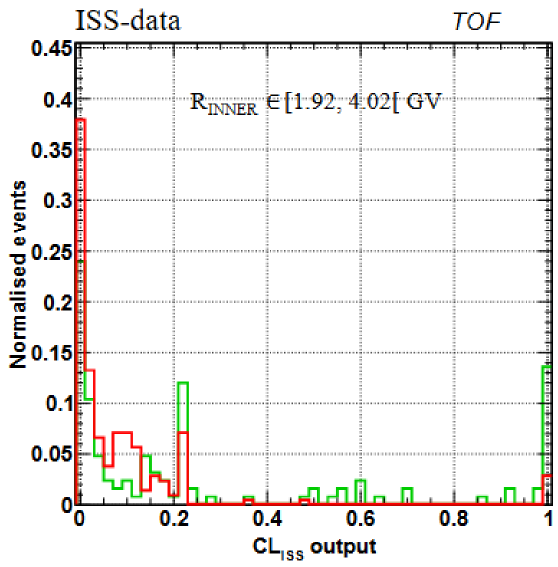
ISS-data
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
CL_{ISS} output



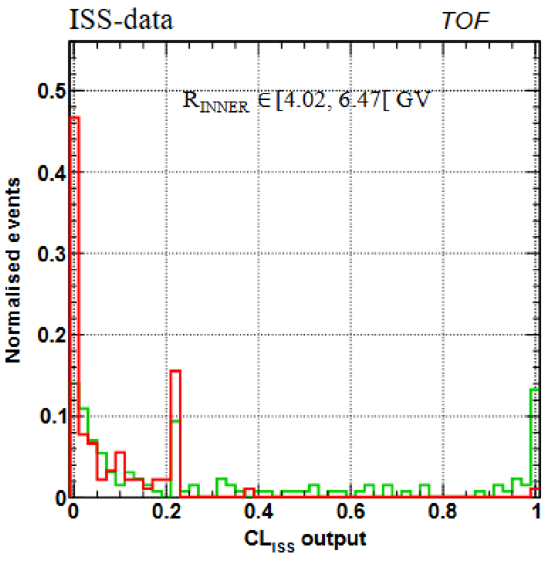
^{4}He MC simulation
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised weights
CL_{MC} output



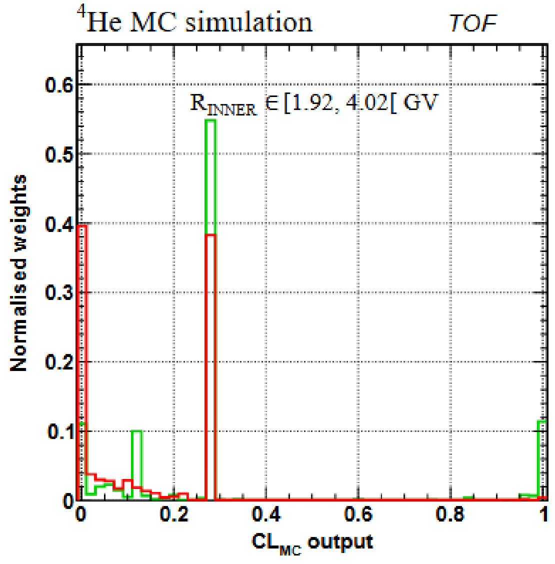
^{4}He MC simulation
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
CL_{MC} output



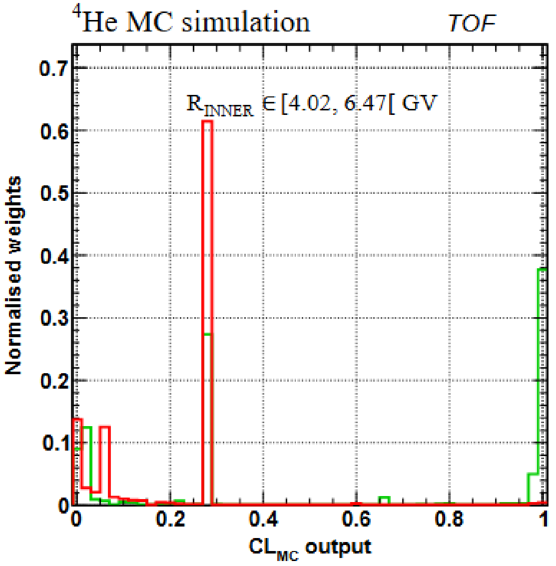
ISS-data
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised events
CL_{ISS} output



ISS-data
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
CL_{ISS} output

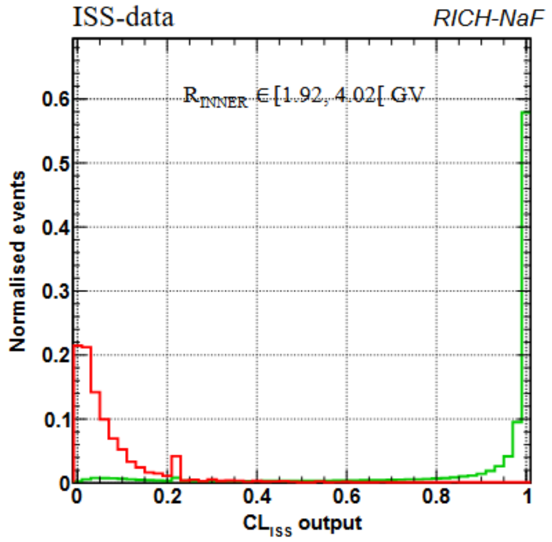


^{4}He MC simulation
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised weights
CL_{MC} output

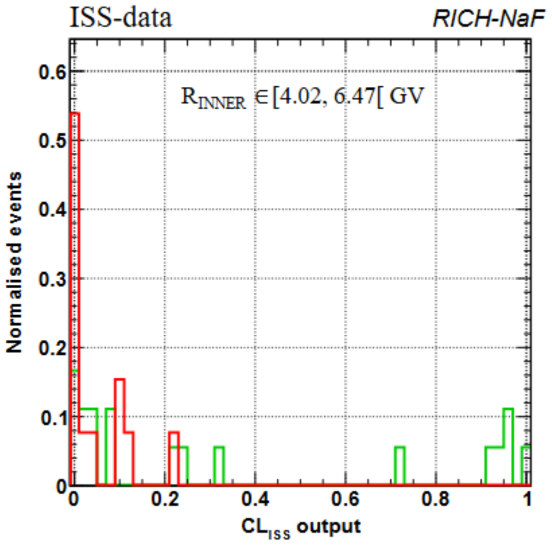


^{4}He MC simulation
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
CL_{MC} output

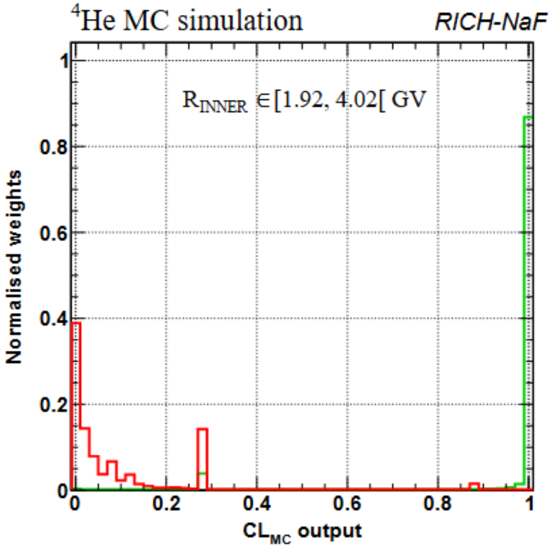
Scores NaF 1



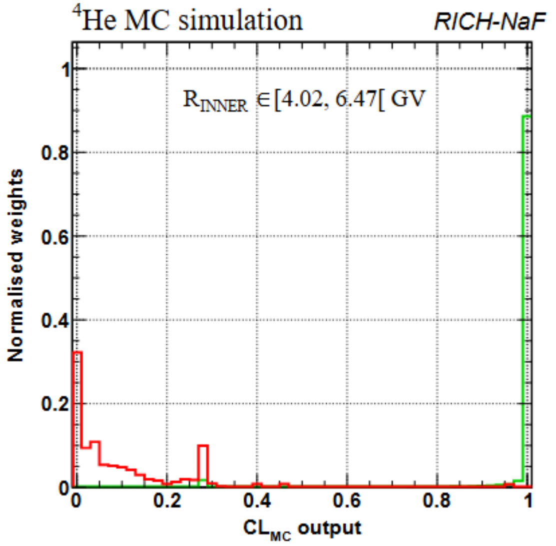
ISS-data
RICH-NaF
R_INNER ∈[1.92, 4.02[GV
Normalised events
CL_ISS output



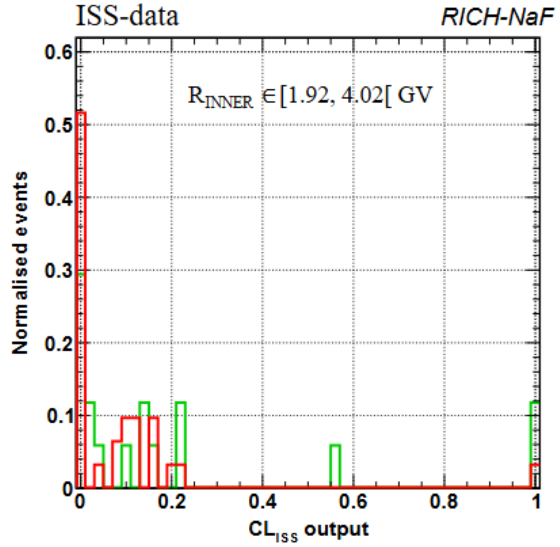
ISS-data
RICH-NaF
R_INNER ∈[4.02, 6.47[GV
Normalised events
CL_ISS output



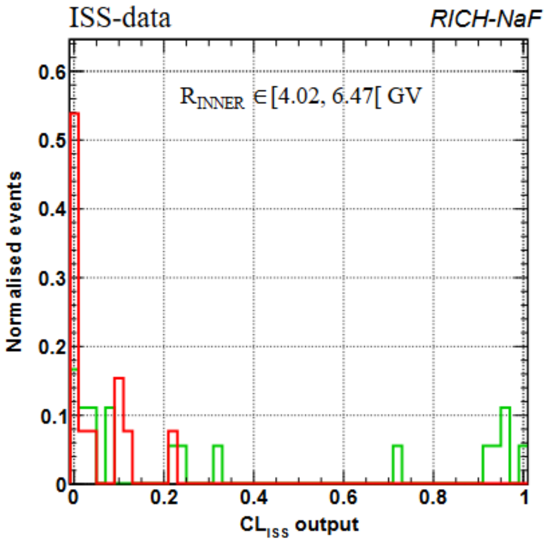
4He MC simulation
RICH-NaF
R_INNER ∈[1.92, 4.02[GV
Normalised weights
CL_MC output



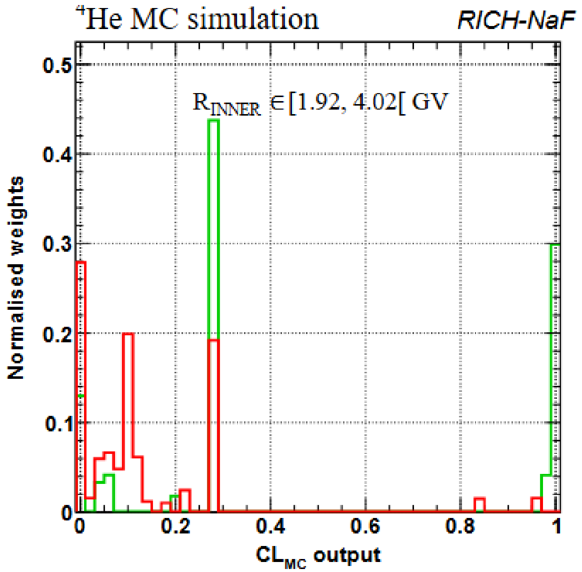
4He MC simulation
RICH-NaF
R_INNER ∈[4.02, 6.47[GV
Normalised weights
CL_MC output



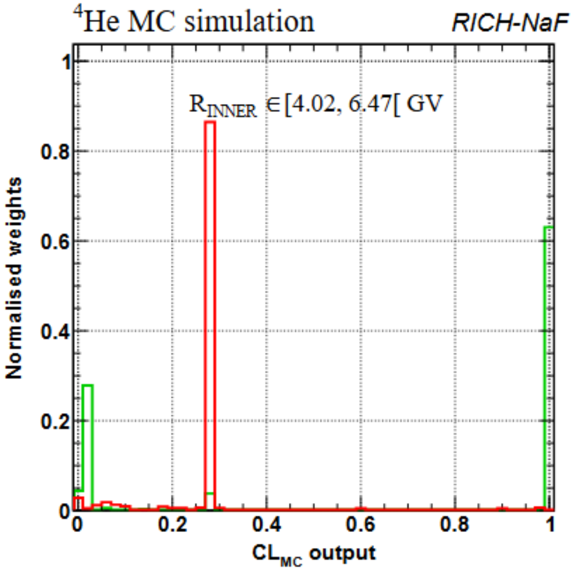
ISS-data
RICH-NaF
R_INNER ∈[1.92, 4.02[GV
Normalised events
CL_ISS output



ISS-data
RICH-NaF
R_INNER ∈[4.02, 6.47[GV
Normalised events
CL_ISS output

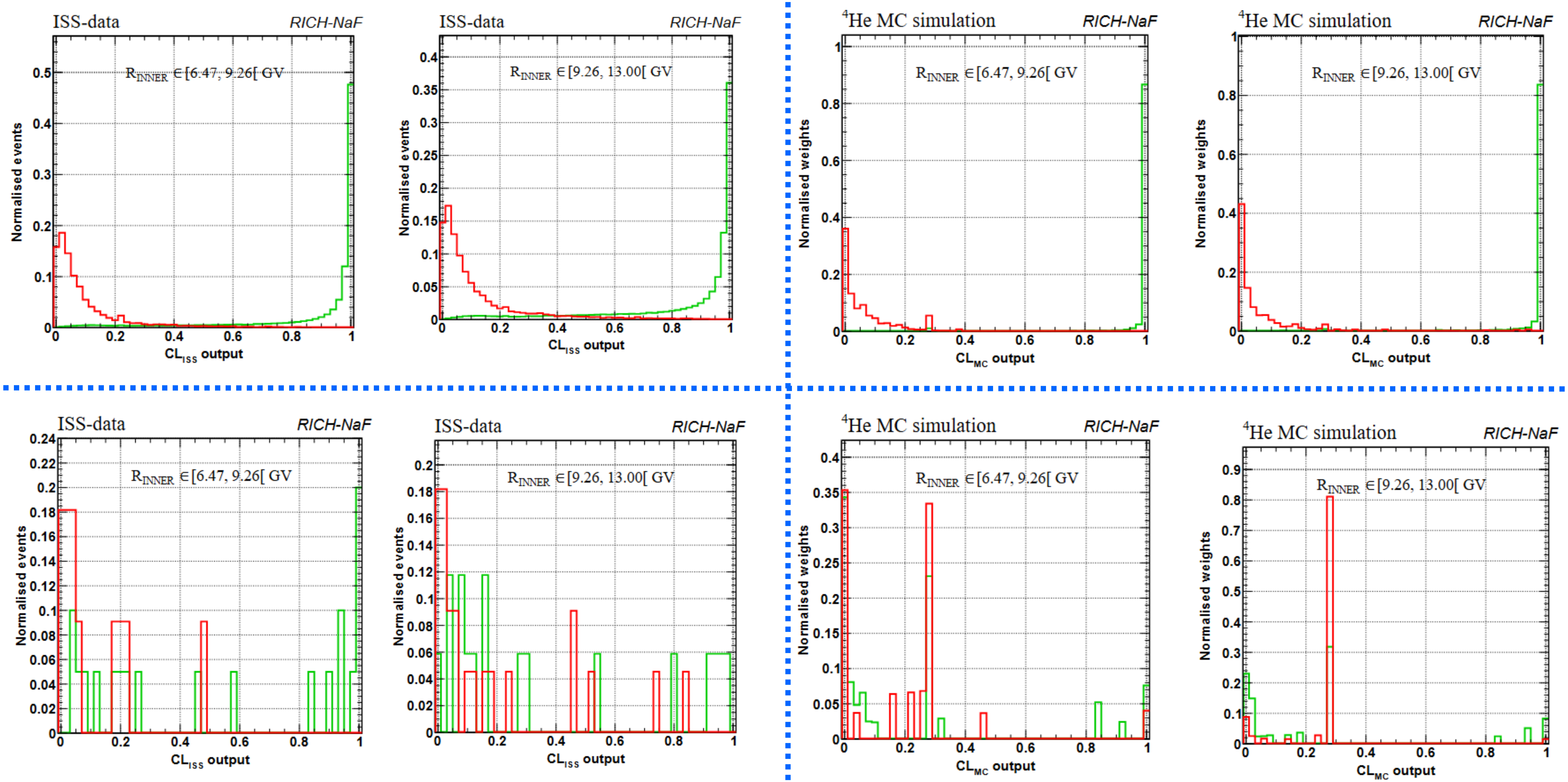


4He MC simulation
RICH-NaF
R_INNER ∈[1.92, 4.02[GV
Normalised weights
CL_MC output



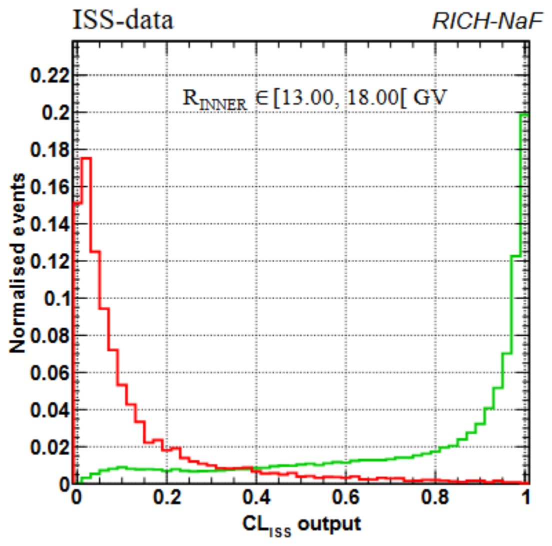
4He MC simulation
RICH-NaF
R_INNER ∈[4.02, 6.47[GV
Normalised weights
CL_MC output

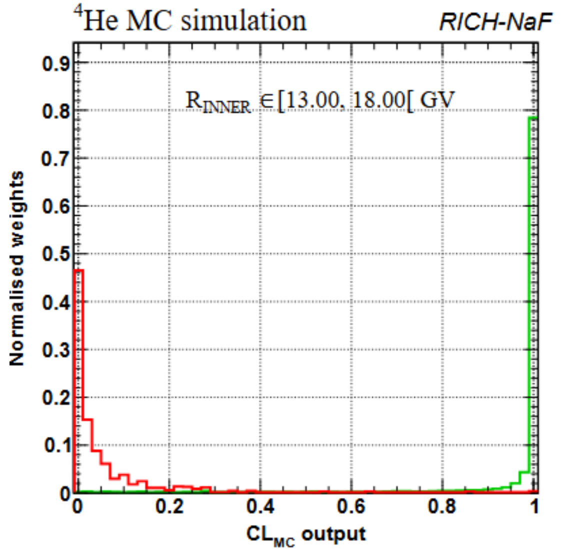
Scores NaF 2

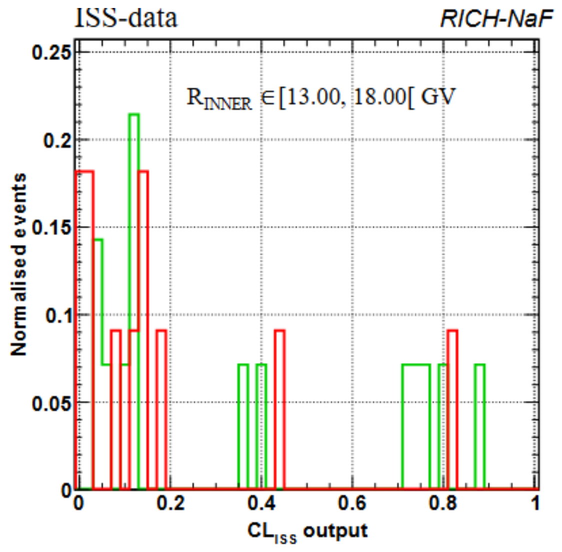


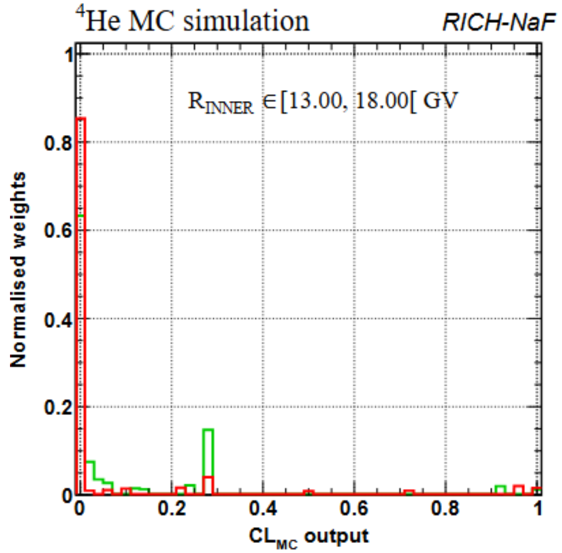
ISS-data
RICH-NaF
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised events
CL_{ISS} output
ISS-data
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised events
CL_{ISS} output
^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised weights
CL_{MC} output
^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised weights
CL_{MC} output
ISS-data
RICH-NaF
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised events
CL_{ISS} output
ISS-data
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised events
CL_{ISS} output
^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised weights
CL_{MC} output
^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised weights
CL_{MC} output

Scores NaF 3

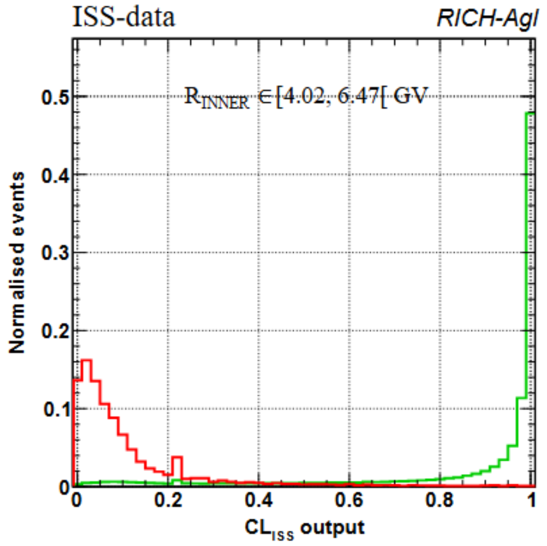

ISS-data
RICH-NaF
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised events
CL_{ISS} output

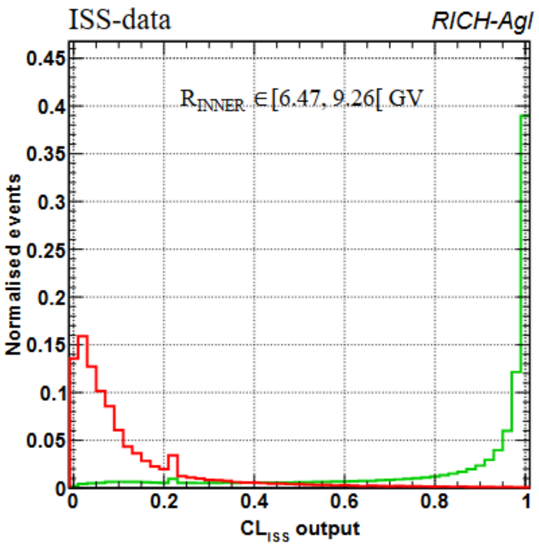

^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised weights
CL_{MC} output

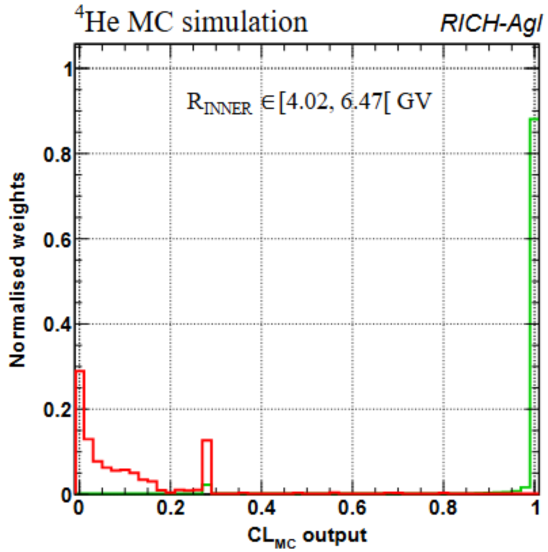

ISS-data
RICH-NaF
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised events
CL_{ISS} output

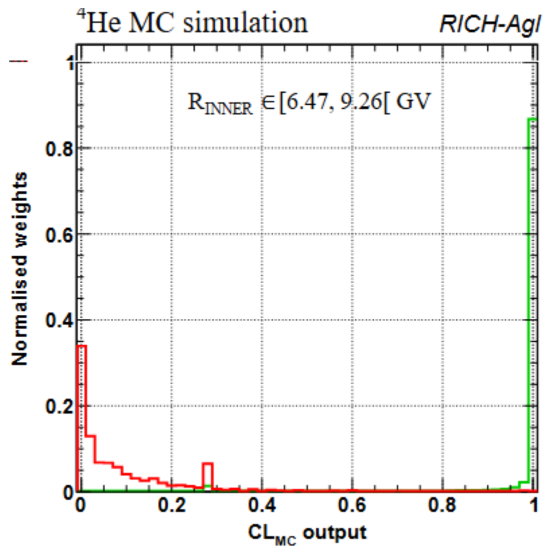

^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised weights
CL_{MC} output

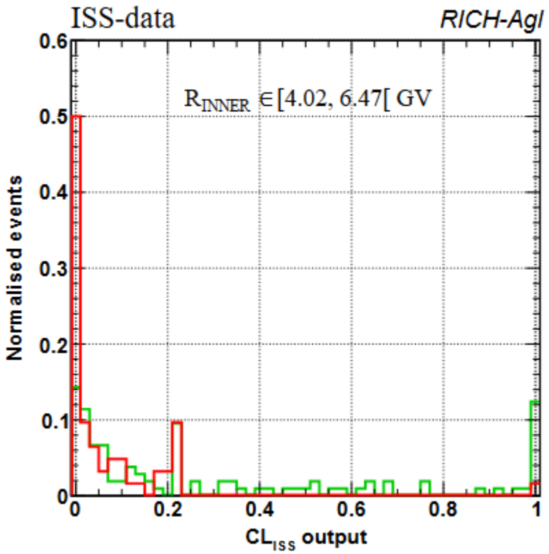
Scores AGL 1

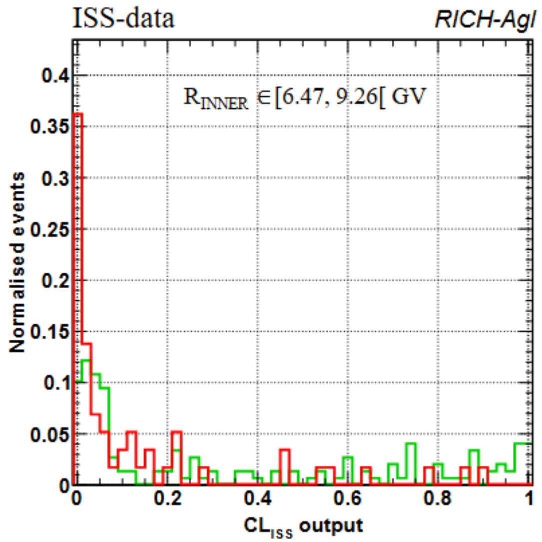

ISS-data
RICH-Agl
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
CL_{ISS} output

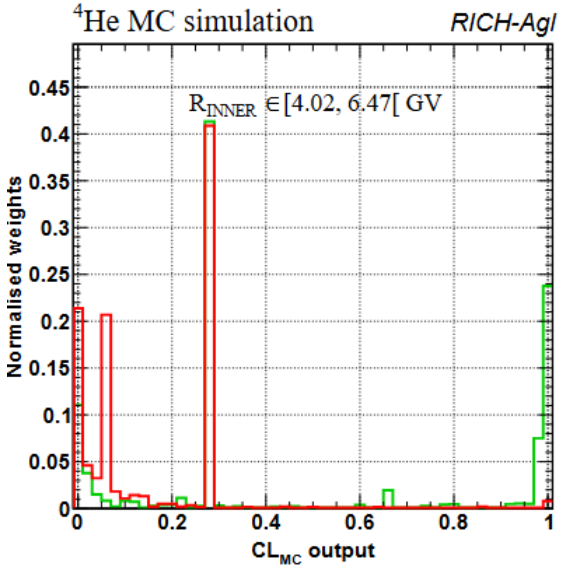

ISS-data
RICH-Agl
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised events
CL_{ISS} output

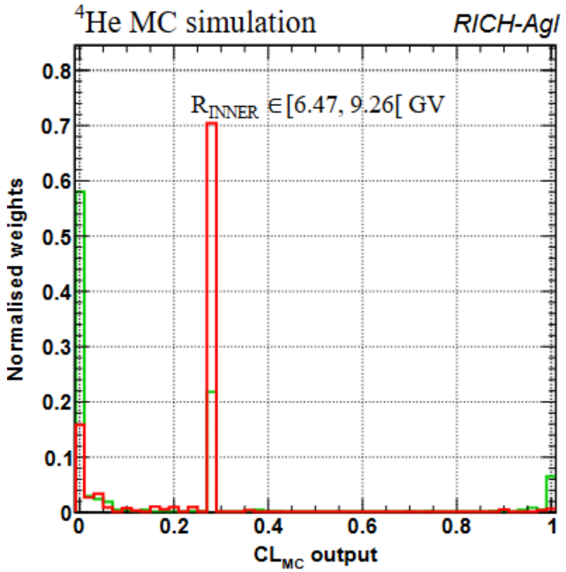

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
CL_{MC} output


^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised weights
CL_{MC} output

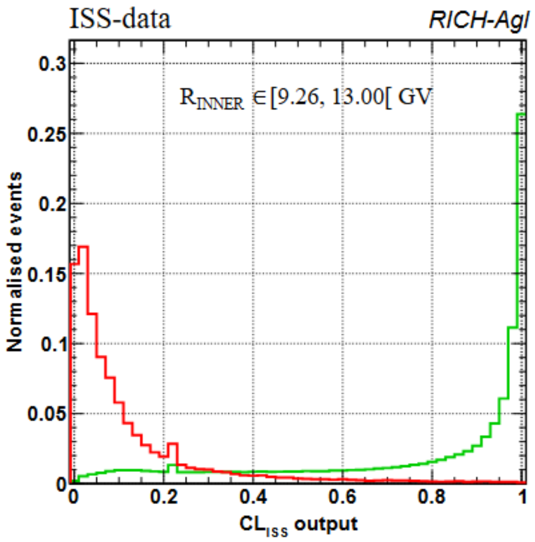

ISS-data
RICH-Agl
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
CL_{ISS} output


ISS-data
RICH-Agl
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised events
CL_{ISS} output

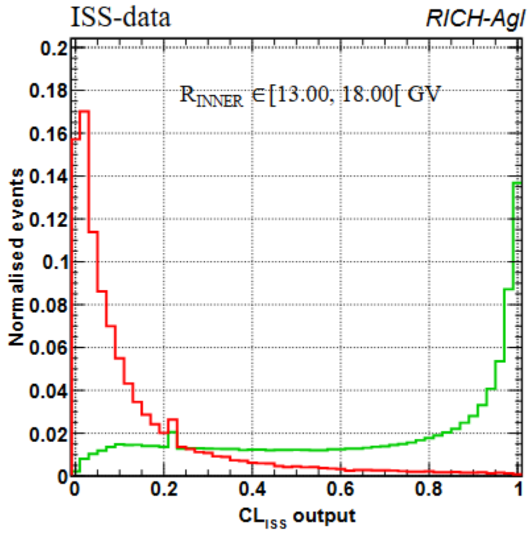

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
CL_{MC} output


^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised weights
CL_{MC} output

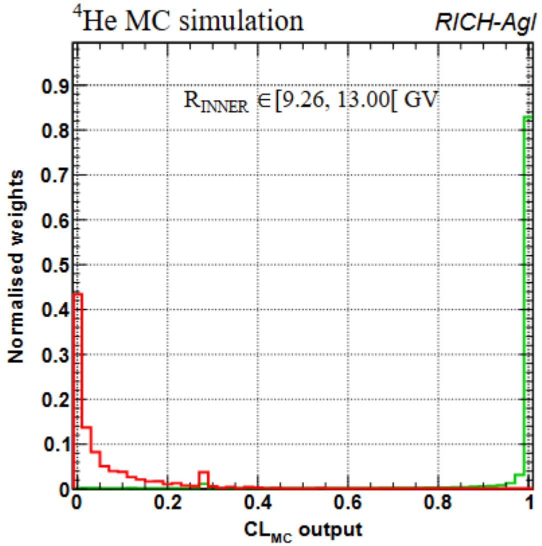
Scores AGL 2



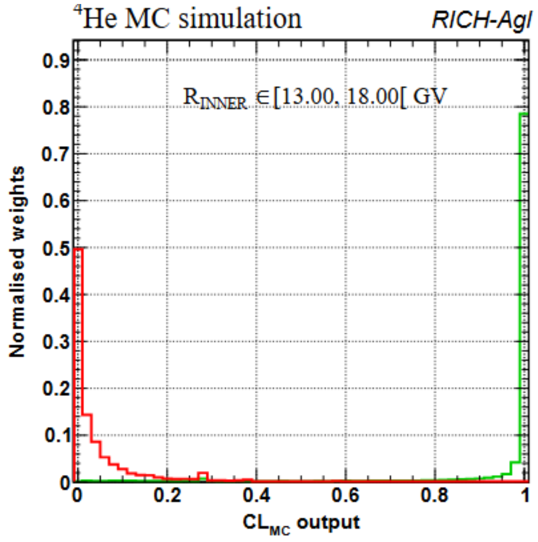
ISS-data
RICH-Agl
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised events
CL_{ISS} output



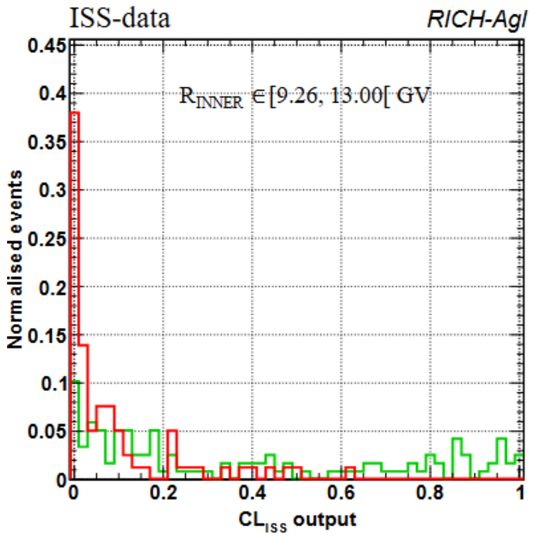
ISS-data
RICH-Agl
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised events
CL_{ISS} output



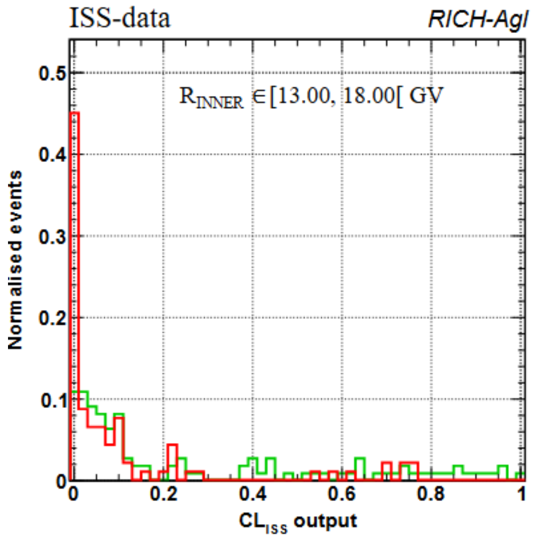
^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised weights
CL_{MC} output



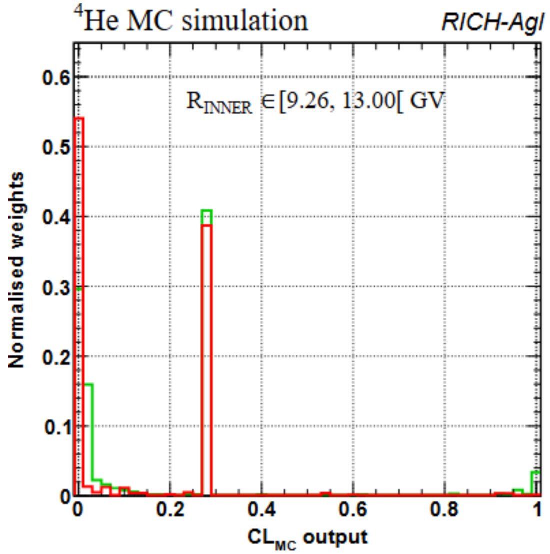
^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised weights
CL_{MC} output



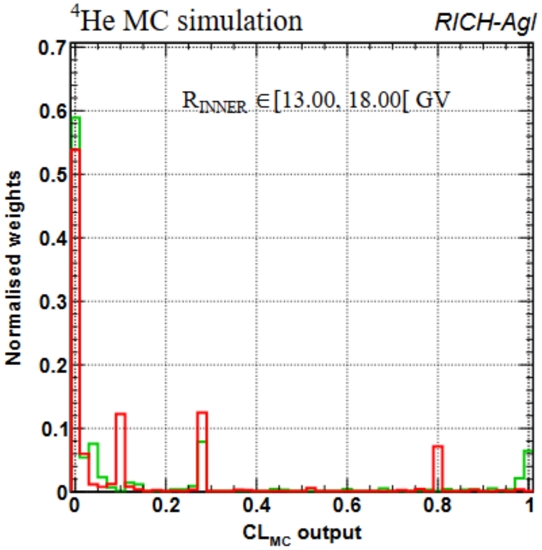
ISS-data
RICH-Agl
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised events
CL_{ISS} output



ISS-data
RICH-Agl
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised events
CL_{ISS} output

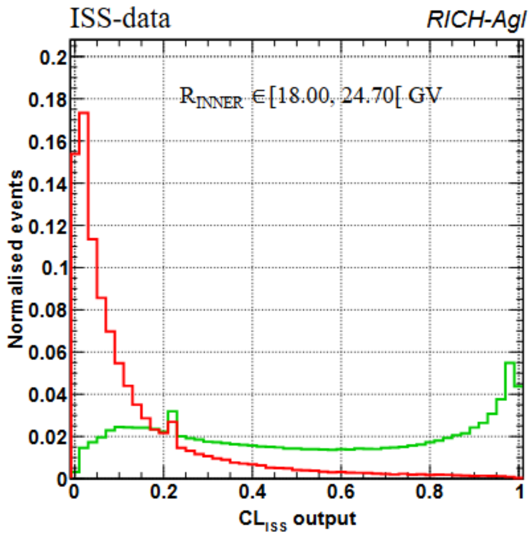


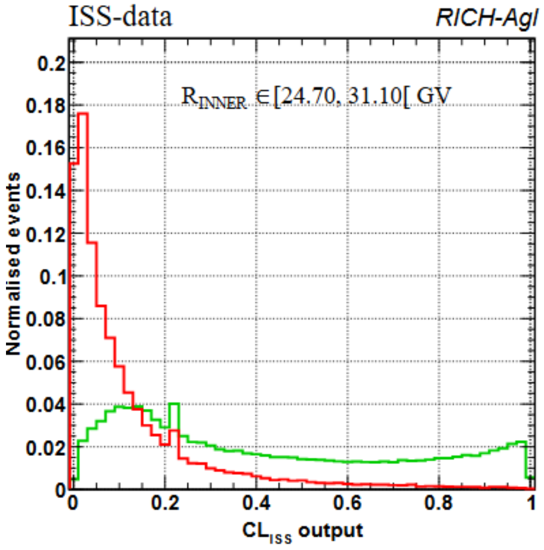
^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised weights
CL_{MC} output

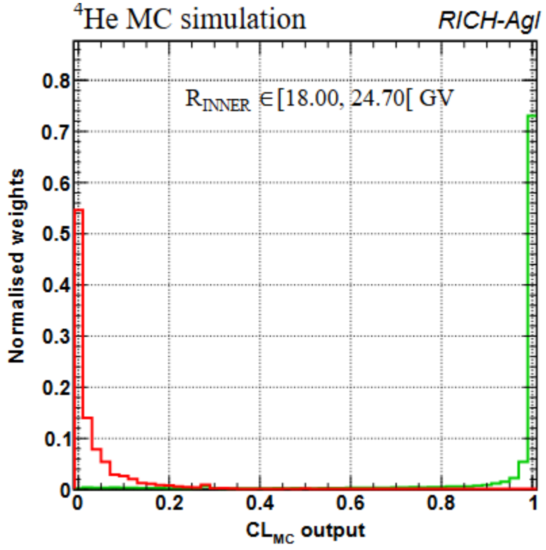


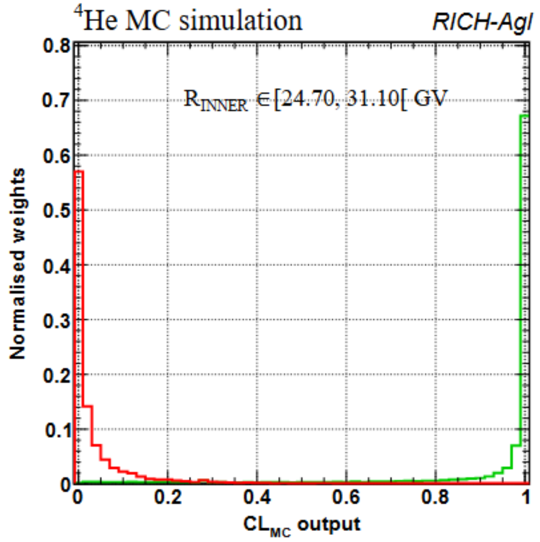
^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [13.00, 18.00[$ GV
Normalised weights
CL_{MC} output

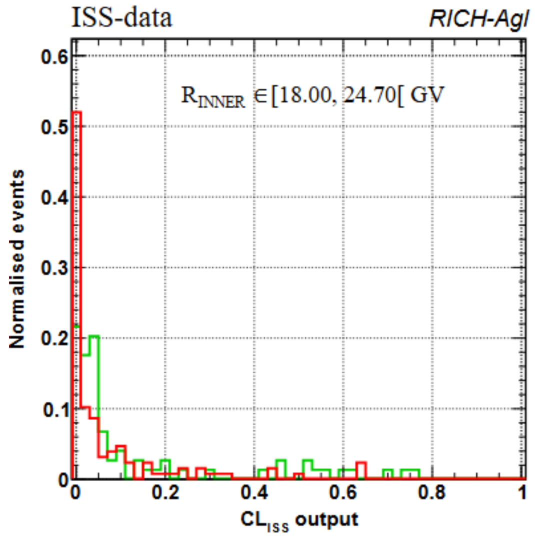
Scores AGL 3

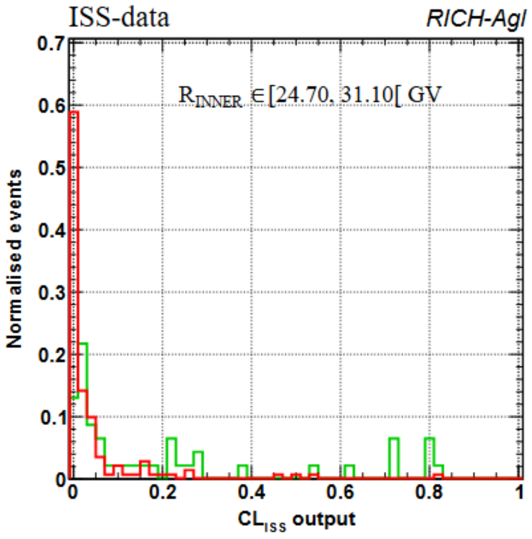

ISS-data
RICH-Agl
$R_{INNER} \in [18.00, 24.70[$ GV
Normalised events
CL_{ISS} output

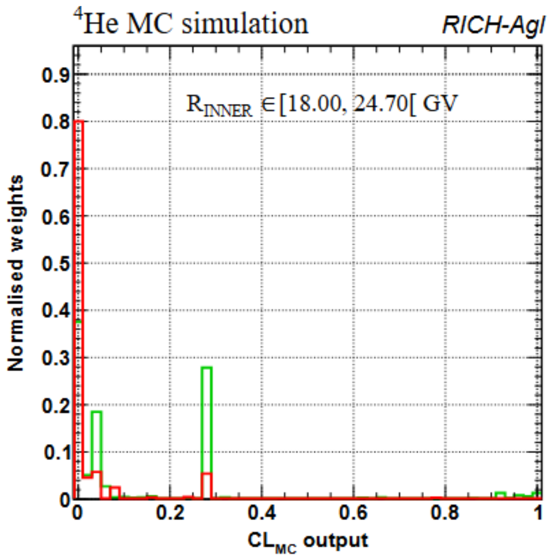

ISS-data
RICH-Agl
$R_{INNER} \in [24.70, 31.10[$ GV
Normalised events
CL_{ISS} output

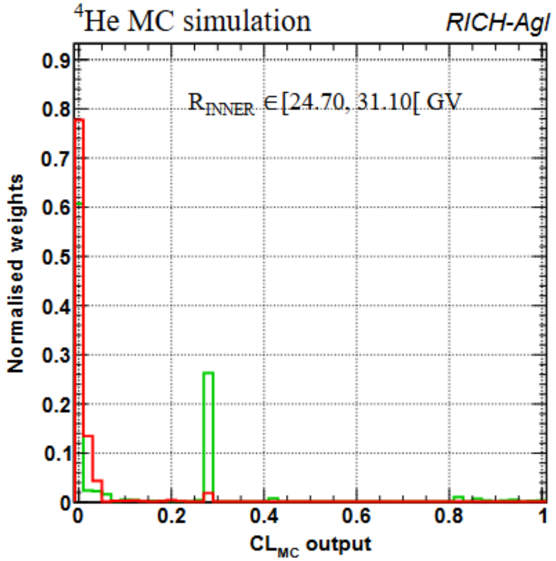

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [18.00, 24.70[$ GV
Normalised weights
CL_{MC} output


^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [24.70, 31.10[$ GV
Normalised weights
CL_{MC} output

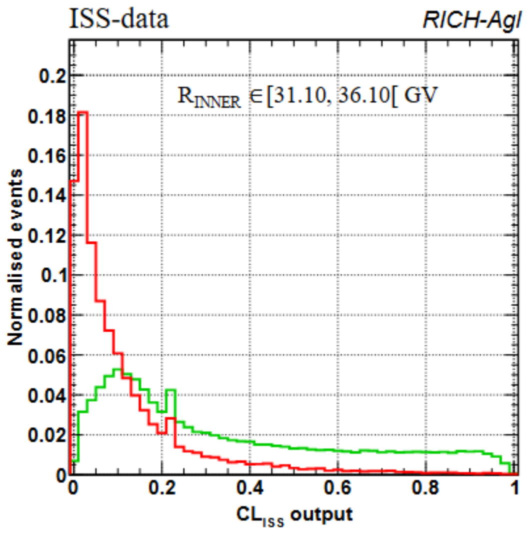

ISS-data
RICH-Agl
$R_{INNER} \in [18.00, 24.70[$ GV
Normalised events
CL_{ISS} output

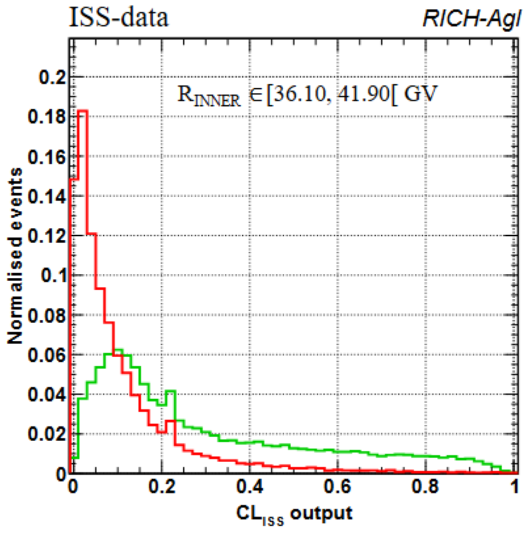

ISS-data
RICH-Agl
$R_{INNER} \in [24.70, 31.10[$ GV
Normalised events
CL_{ISS} output

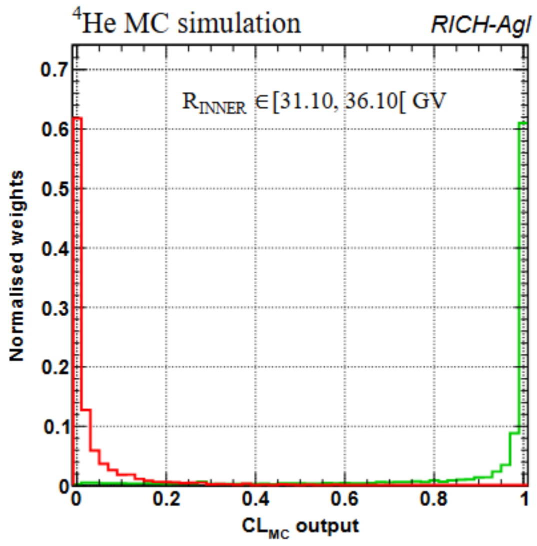

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [18.00, 24.70[$ GV
Normalised weights
CL_{MC} output

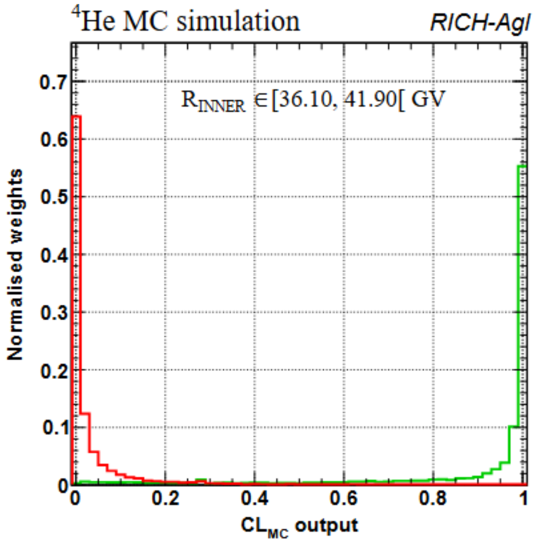

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [24.70, 31.10[$ GV
Normalised weights
CL_{MC} output

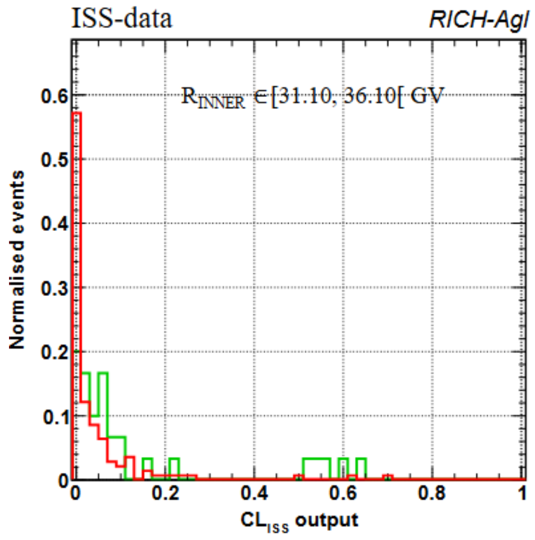
Scores AGL 4

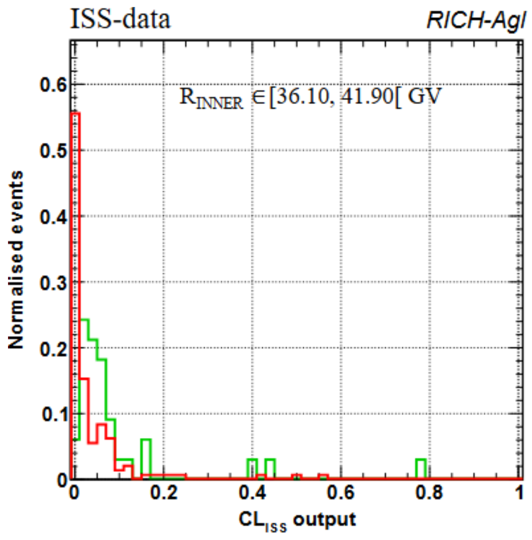

ISS-data
RICH-Agl
$R_{INNER} \in [31.10, 36.10[$ GV
Normalised events
CL_{ISS} output

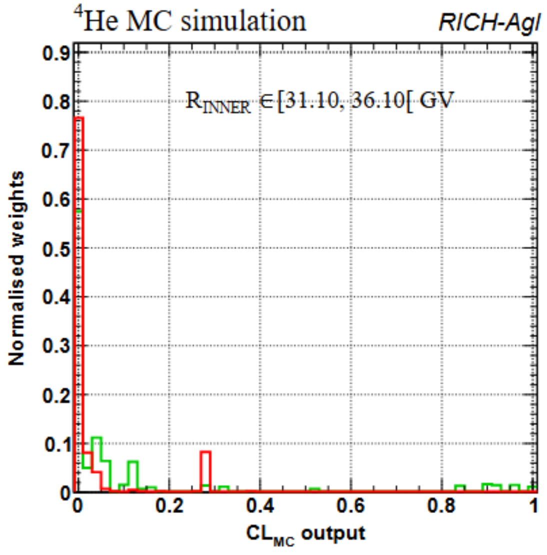

ISS-data
RICH-Agl
$R_{INNER} \in [36.10, 41.90[$ GV
Normalised events
CL_{ISS} output

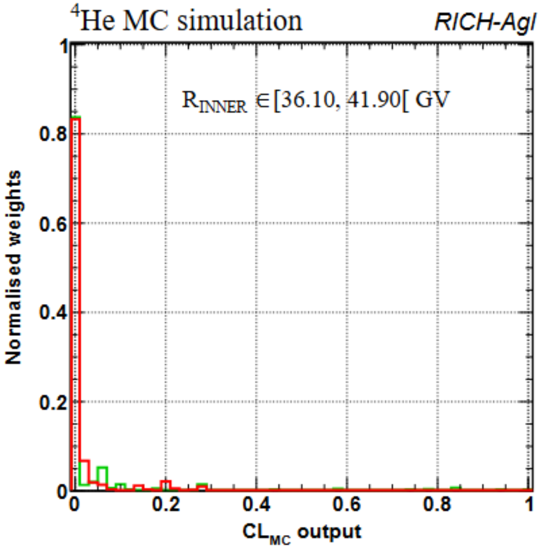

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [31.10, 36.10[$ GV
Normalised weights
CL_{MC} output


^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [36.10, 41.90[$ GV
Normalised weights
CL_{MC} output

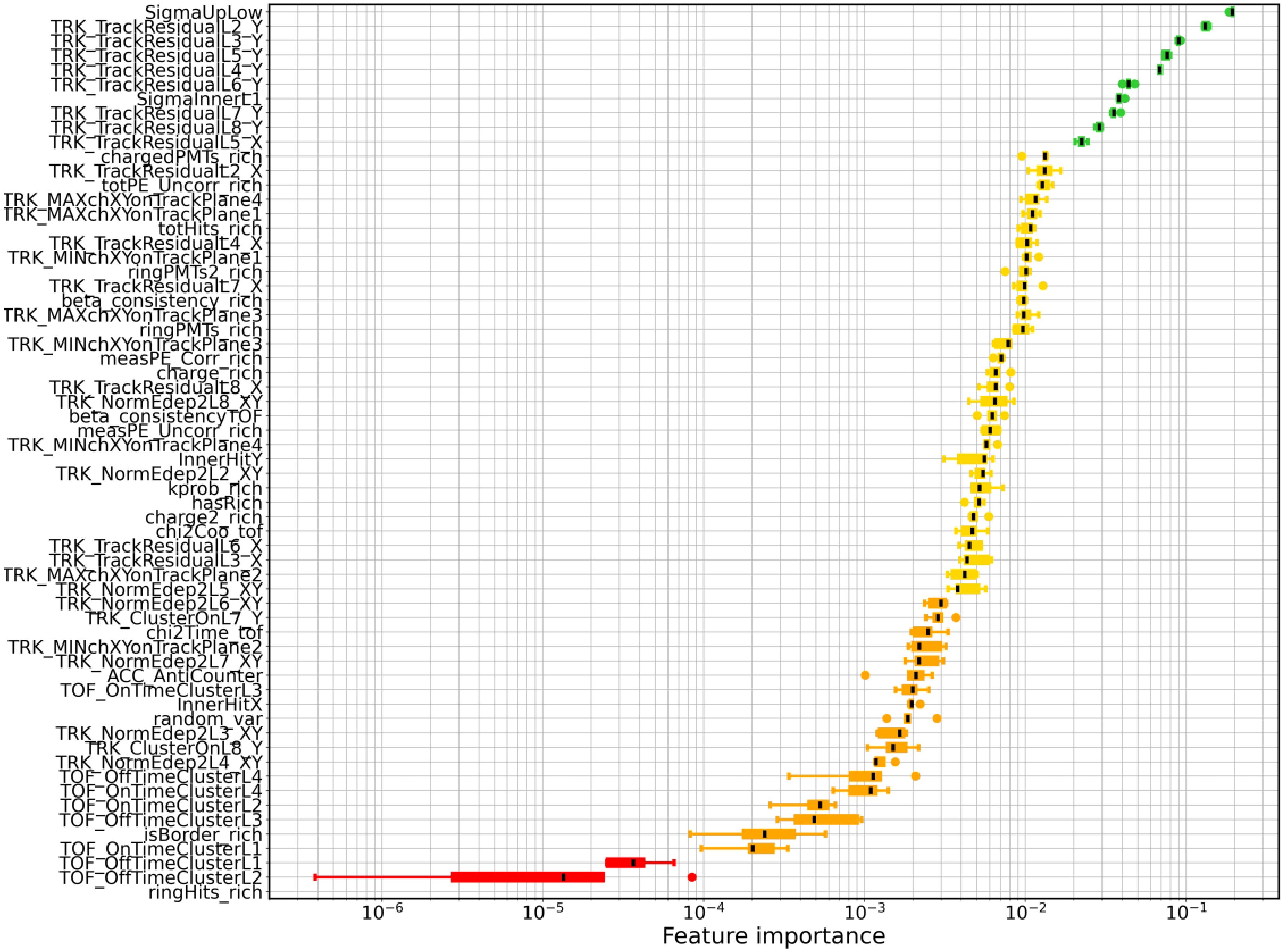

ISS-data
RICH-Agl
$R_{INNER} \in [31.10, 36.10[$ GV
Normalised events
CL_{ISS} output


ISS-data
RICH-Agl
$R_{INNER} \in [36.10, 41.90[$ GV
Normalised events
CL_{ISS} output

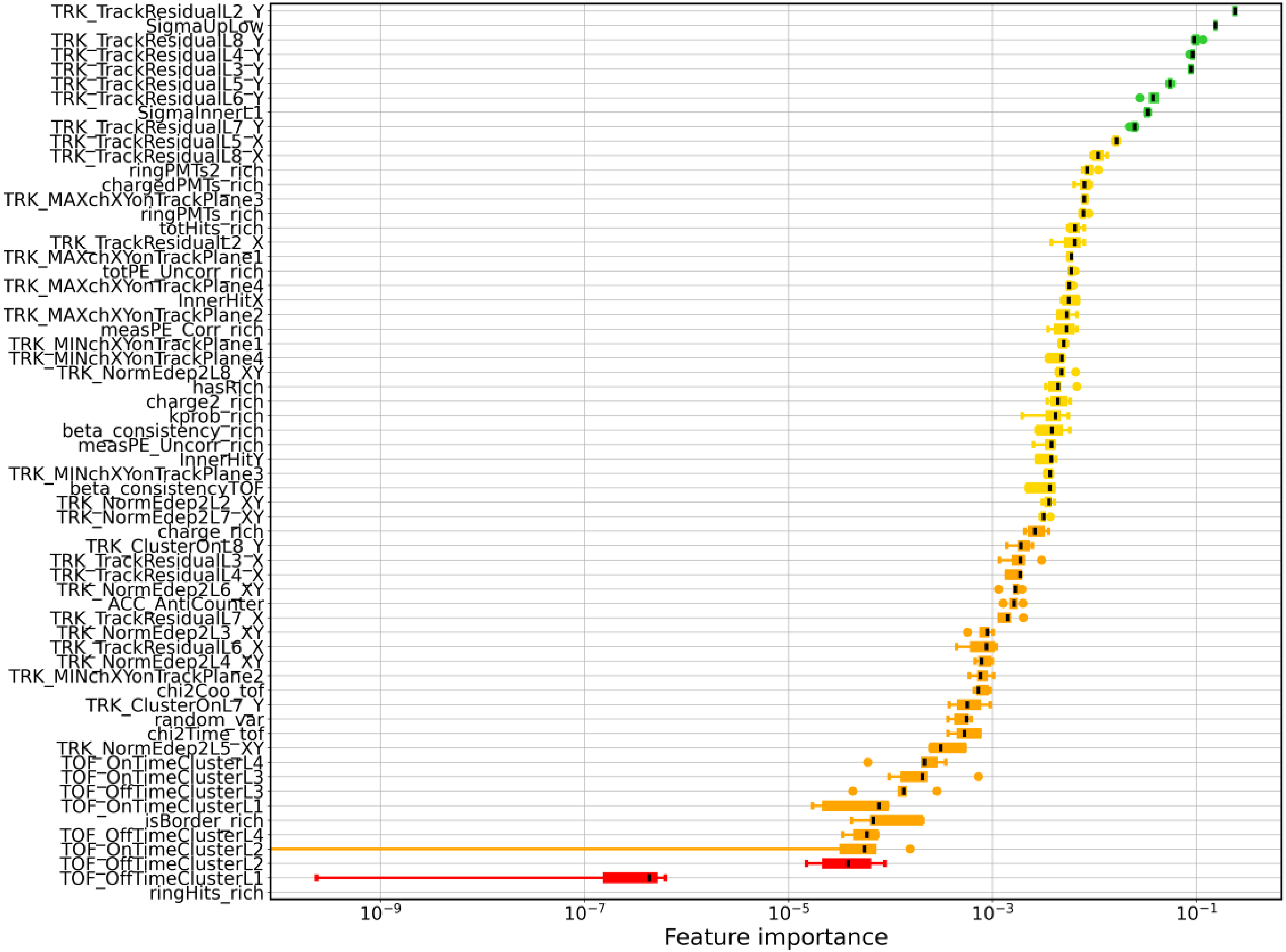

^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [31.10, 36.10[$ GV
Normalised weights
CL_{MC} output


^{4}He MC simulation
RICH-Agl
$R_{INNER} \in [36.10, 41.90[$ GV
Normalised weights
CL_{MC} output

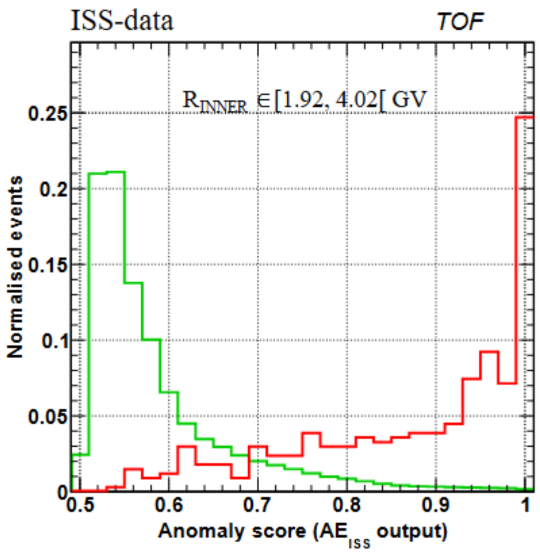
Feature Ranking
AE_{ISS}

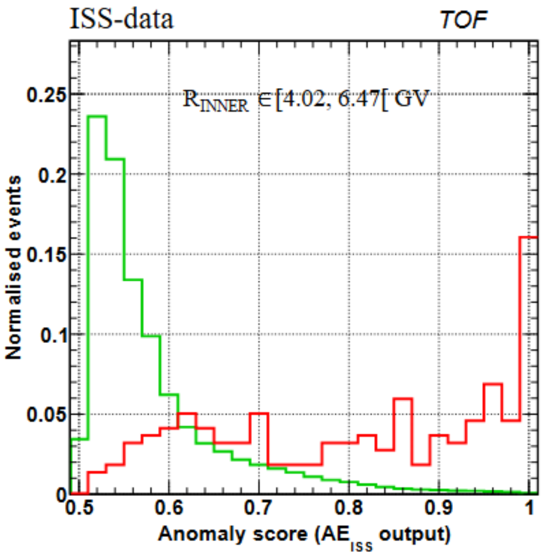


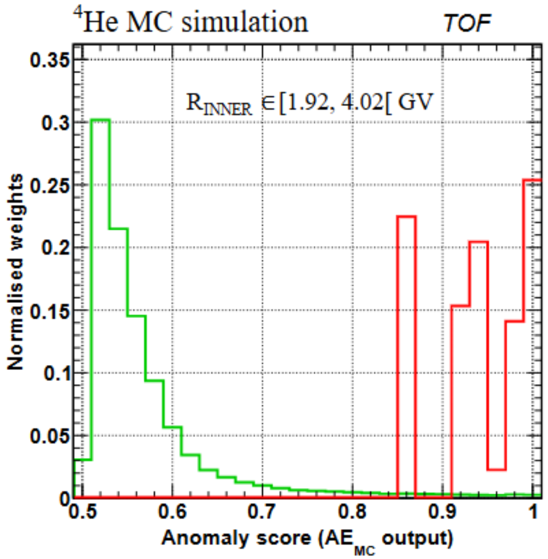
Feature Ranking
AE_{MC}

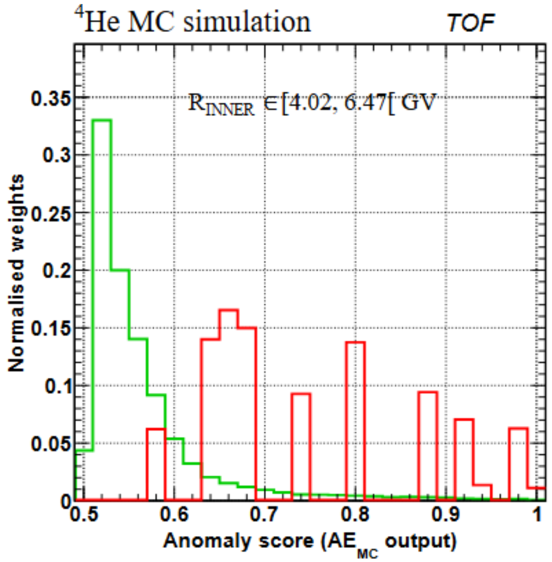


Scores TOF

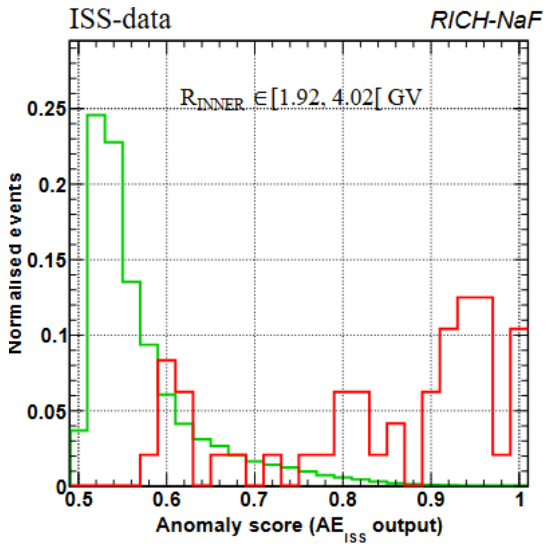

ISS-data
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised events
Anomaly score (AE_{ISS} output)

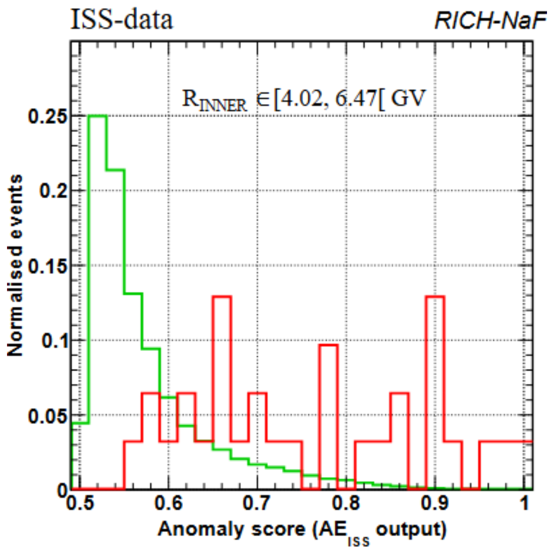

ISS-data
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
Anomaly score (AE_{ISS} output)

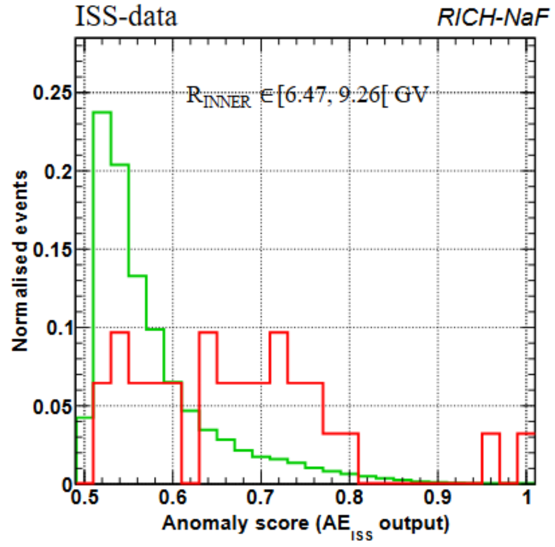

^{4}He MC simulation
TOF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised weights
Anomaly score (AE_{MC} output)

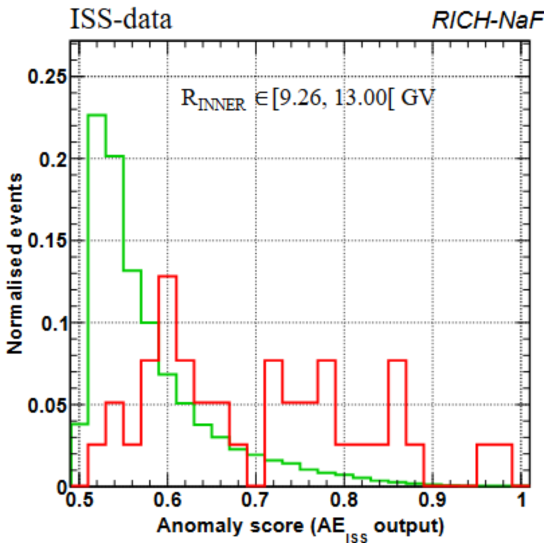

^{4}He MC simulation
TOF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
Anomaly score (AE_{MC} output)

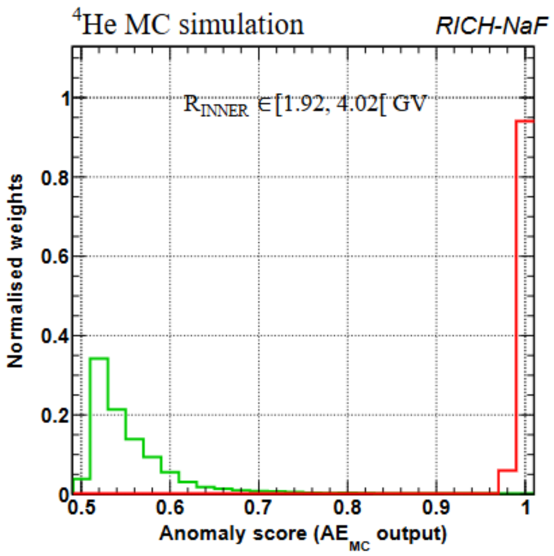
Scores NaF 1

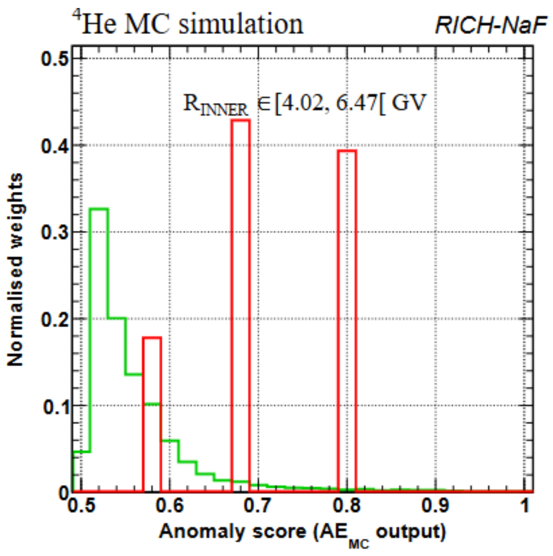

ISS-data
RICH-NaF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised events
Anomaly score (AE_{ISS} output)

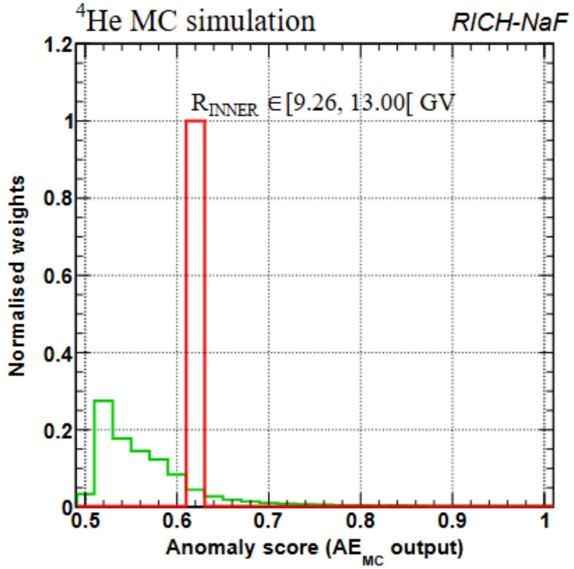

ISS-data
RICH-NaF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised events
Anomaly score (AE_{ISS} output)


ISS-data
RICH-NaF
$R_{INNER} \in [6.47, 9.26[$ GV
Normalised events
Anomaly score (AE_{ISS} output)

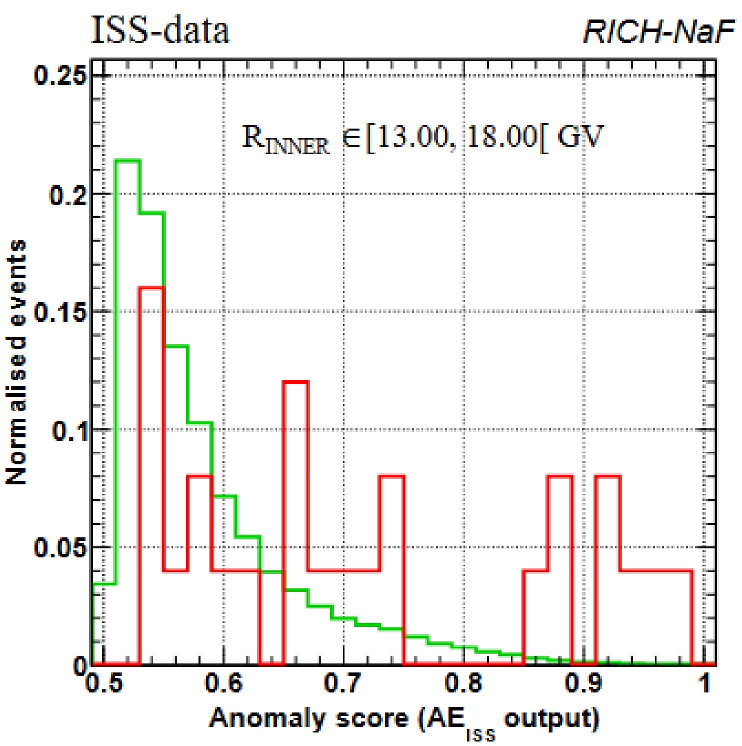

ISS-data
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised events
Anomaly score (AE_{ISS} output)


^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [1.92, 4.02[$ GV
Normalised weights
Anomaly score (AE_{MC} output)

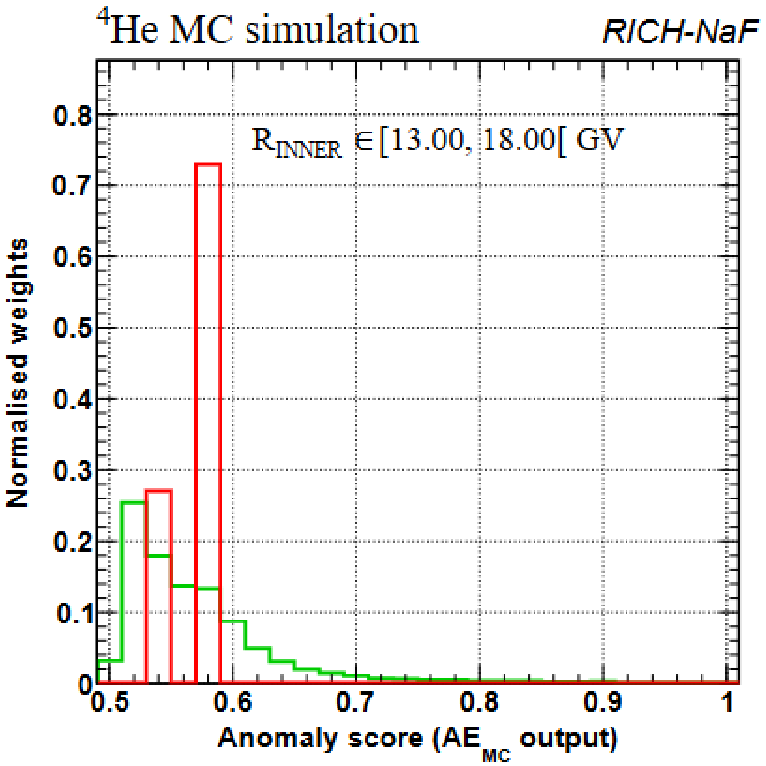

^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [4.02, 6.47[$ GV
Normalised weights
Anomaly score (AE_{MC} output)


^{4}He MC simulation
RICH-NaF
$R_{INNER} \in [9.26, 13.00[$ GV
Normalised weights
Anomaly score (AE_{MC} output)

Scores NaF 2

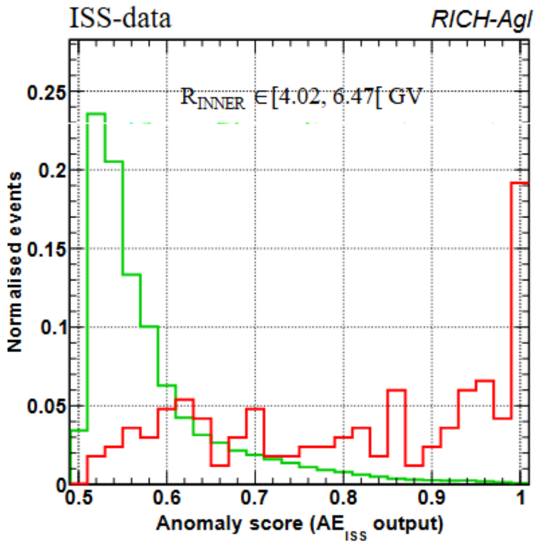


ISS-data
RICH-NaF
R_INNER ∈[13.00, 18.00[GV
Normalised events
Anomaly score (AE_ISS output)
0
0.05
0.1
0.15
0.2
0.25
0.5
0.6
0.7
0.8
0.9
1

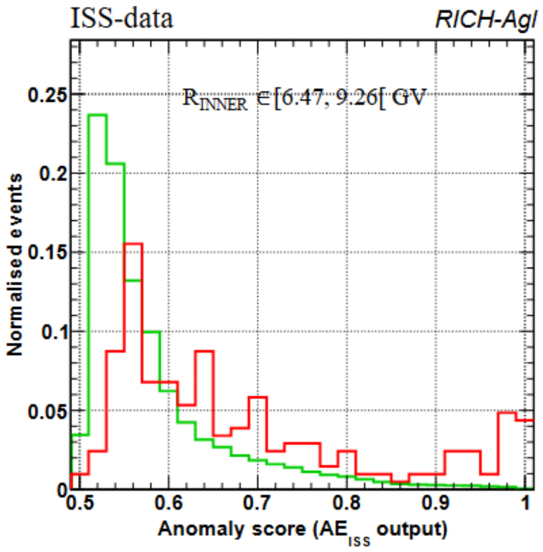


4He MC simulation
RICH-NaF
R_INNER ∈[13.00, 18.00[GV
Normalised weights
Anomaly score (AE_MC output)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.5
0.6
0.7
0.8
0.9
1

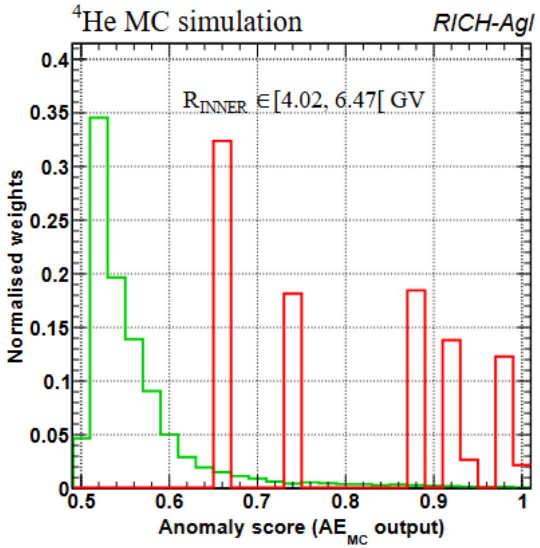
Scores AGL 1



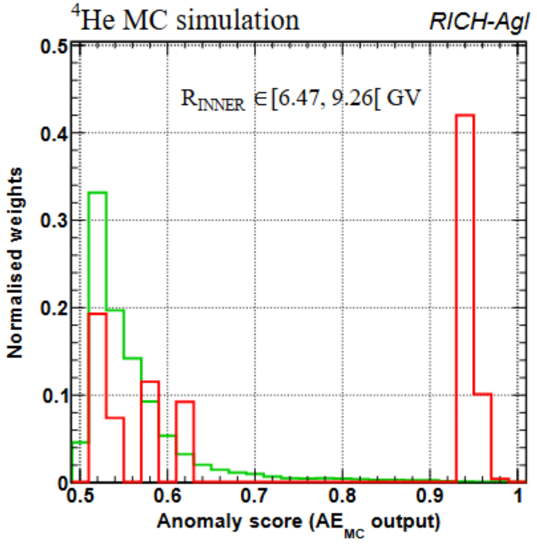
ISS-data
RICH-Agl
R_INNER ∈[4.02, 6.47[GV
Normalised events
Anomaly score (AE_ISS output)



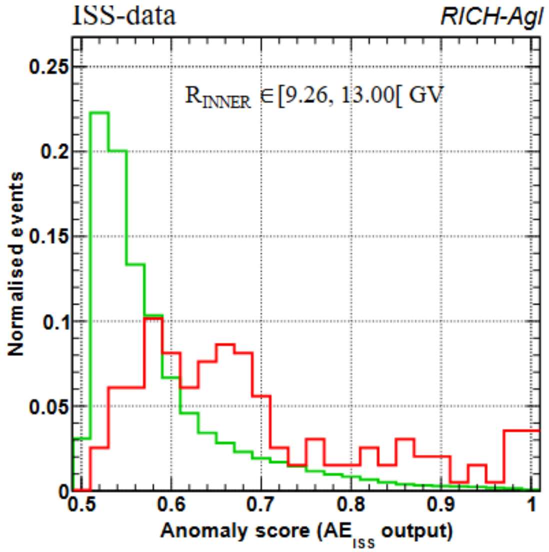
ISS-data
RICH-Agl
R_INNER ∈[6.47, 9.26[GV
Normalised events
Anomaly score (AE_ISS output)



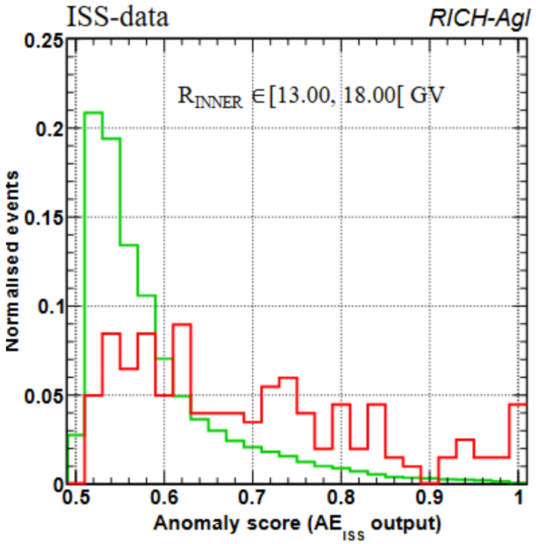
4He MC simulation
RICH-Agl
R_INNER ∈[4.02, 6.47[GV
Normalised weights
Anomaly score (AE_MC output)



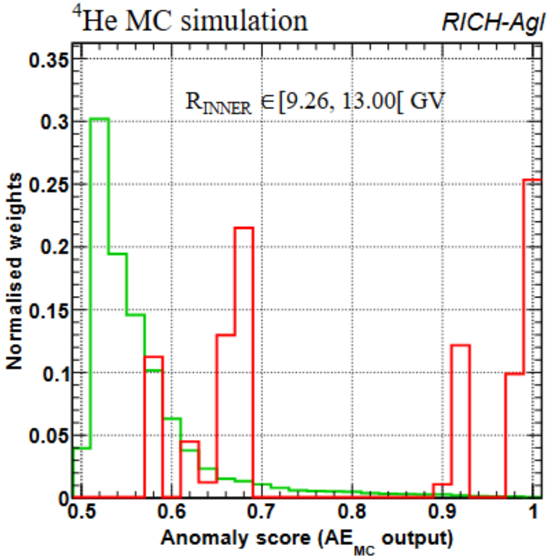
4He MC simulation
RICH-Agl
R_INNER ∈[6.47, 9.26[GV
Normalised weights
Anomaly score (AE_MC output)



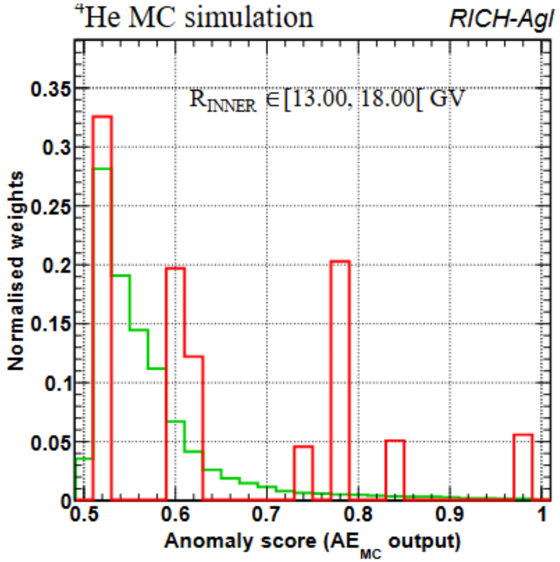
ISS-data
RICH-Agl
R_INNER ∈[9.26, 13.00[GV
Normalised events
Anomaly score (AE_ISS output)



ISS-data
RICH-Agl
R_INNER ∈[13.00, 18.00[GV
Normalised events
Anomaly score (AE_ISS output)

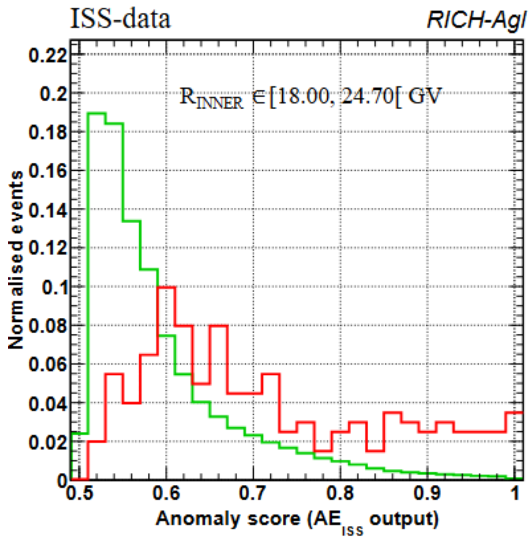


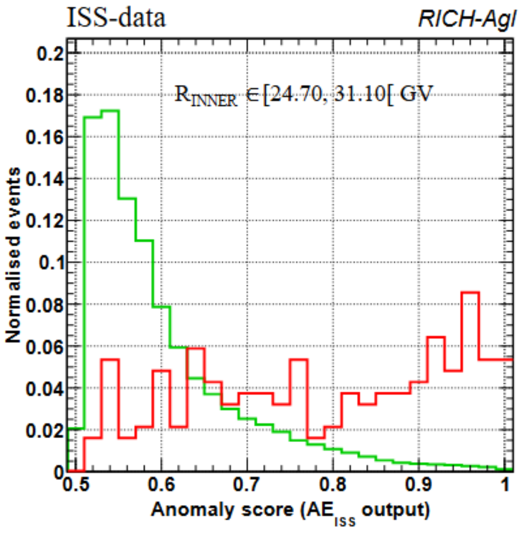
4He MC simulation
RICH-Agl
R_INNER ∈[9.26, 13.00[GV
Normalised weights
Anomaly score (AE_MC output)

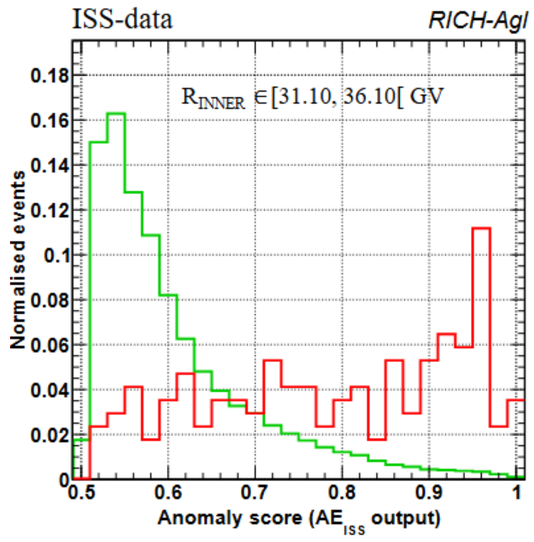


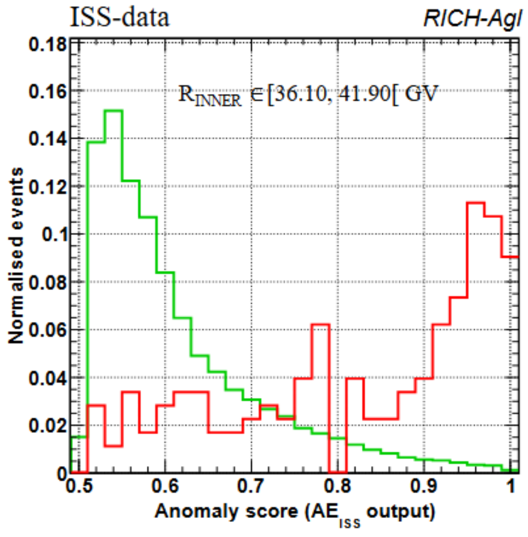
4He MC simulation
RICH-Agl
R_INNER ∈[13.00, 18.00[GV
Normalised weights
Anomaly score (AE_MC output)

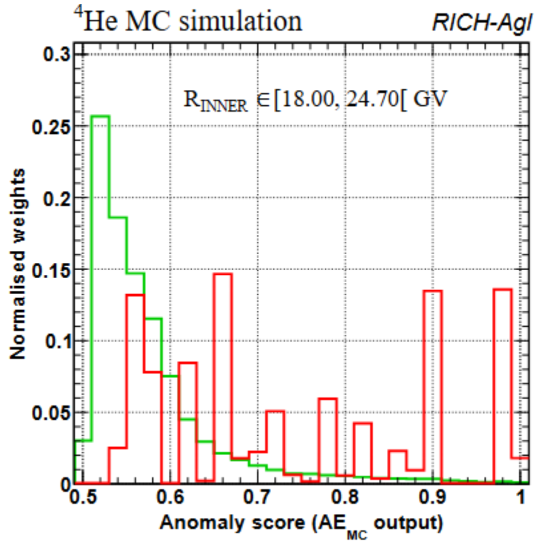
Scores AGL 2

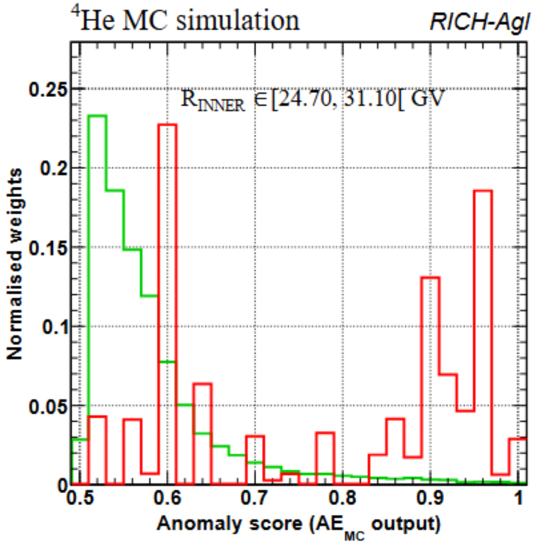

ISS-data
RICH-Agl
R_INNER ∈[18.00, 24.70[GV
Normalised events
Anomaly score (AE_ISS output)

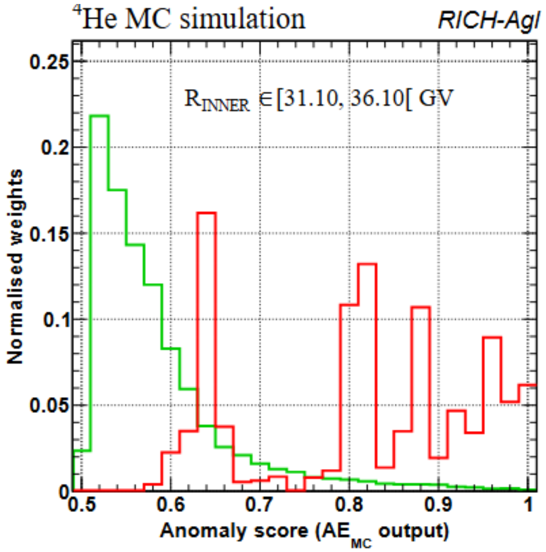

ISS-data
RICH-Agl
R_INNER ∈[24.70, 31.10[GV
Normalised events
Anomaly score (AE_ISS output)

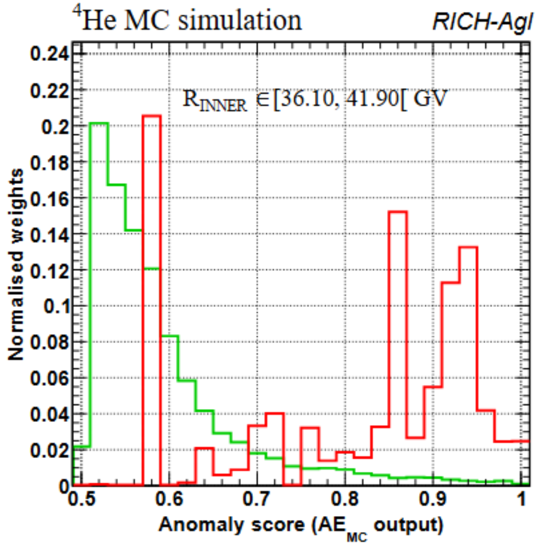

ISS-data
RICH-Agl
R_INNER ∈[31.10, 36.10[GV
Normalised events
Anomaly score (AE_ISS output)


ISS-data
RICH-Agl
R_INNER ∈[36.10, 41.90[GV
Normalised events
Anomaly score (AE_ISS output)

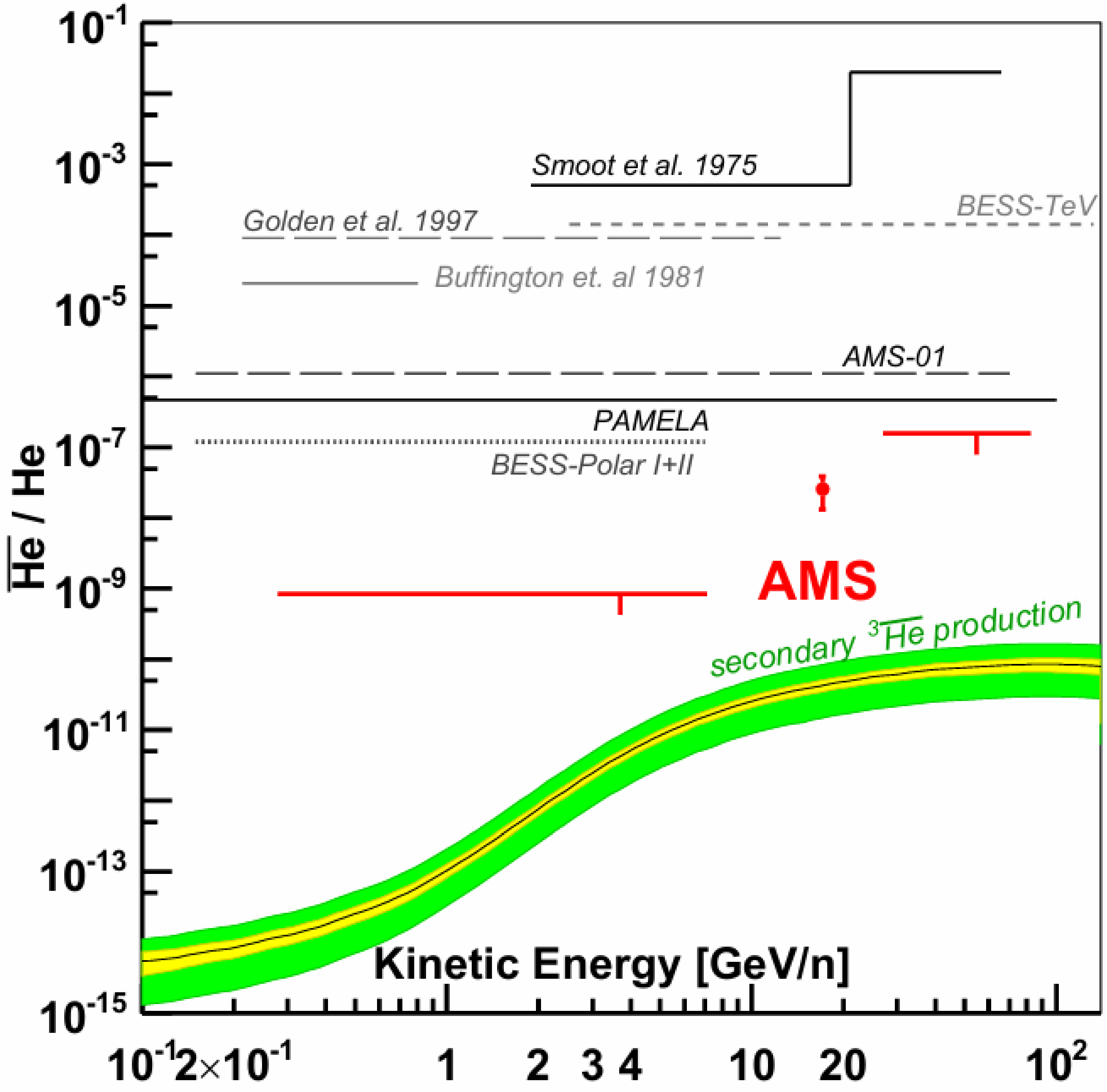

4He MC simulation
RICH-Agl
R_INNER ∈[18.00, 24.70[GV
Normalised weights
Anomaly score (AE_MC output)


4He MC simulation
RICH-Agl
R_INNER ∈[24.70, 31.10[GV
Normalised weights
Anomaly score (AE_MC output)


4He MC simulation
RICH-Agl
R_INNER ∈[31.10, 36.10[GV
Normalised weights
Anomaly score (AE_MC output)


4He MC simulation
RICH-Agl
R_INNER ∈[36.10, 41.90[GV
Normalised weights
Anomaly score (AE_MC output)

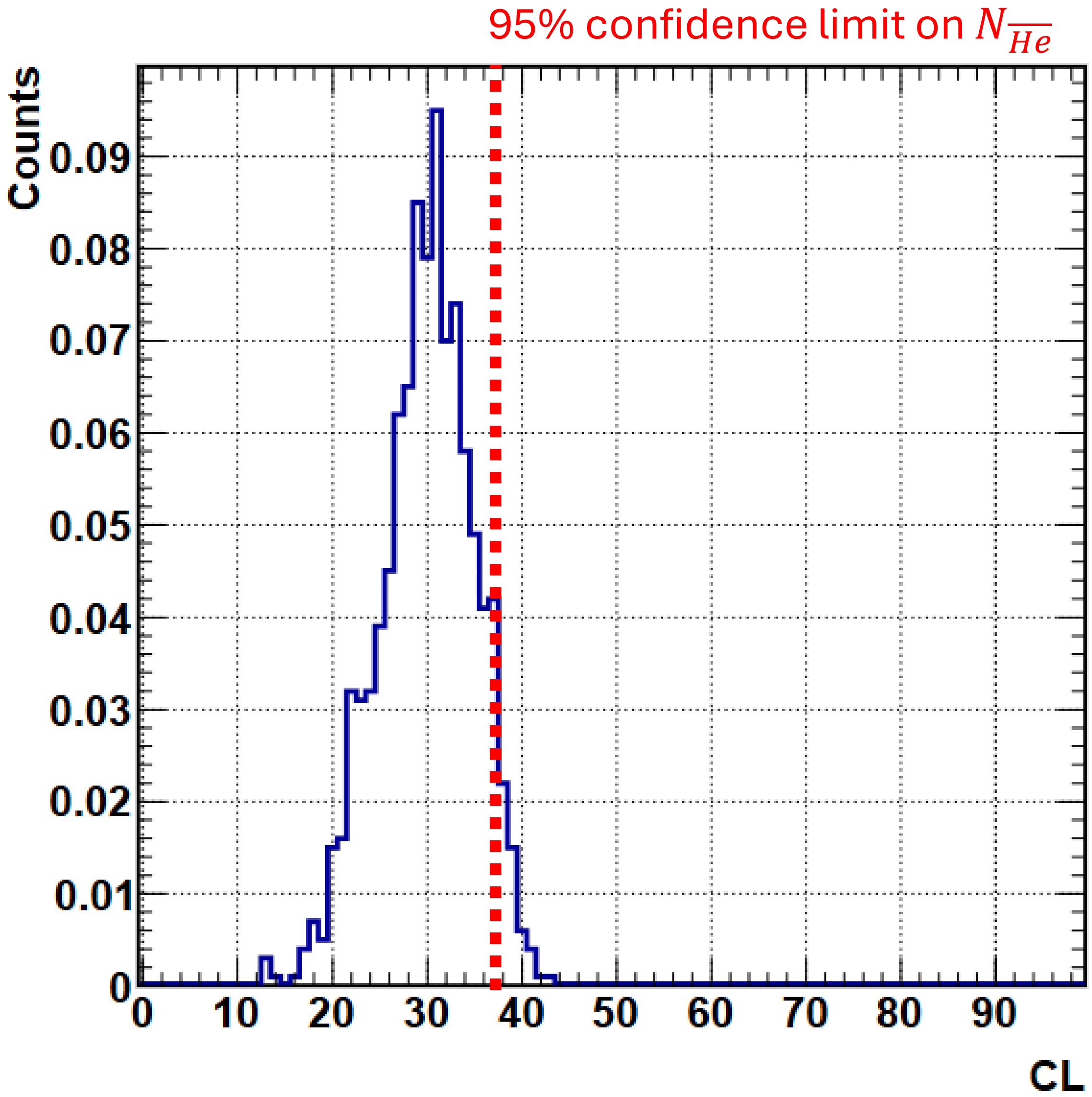
V. Choutko results (GM 06.2025)



Confidence limit estimate

- Identify $\overline{He}$ candidates:
  - additional cut on the reconstructed mass $m_{REC} \in [2,5]$ GeV/c^2
- Number of survivors in data (n_{ISS}) and MC ($\mathrm{n_{MC}}$), integrated in R.
- Monte Carlo toy procedure:
  - Uncertainty on background (from MC) → Poissonian (F_{bkg}) with $\mu = n_{MC}$
  - Extract 1000 times from $F_{bkg} \rightarrow n_{bkg}^{i}$
    - Assume n_{bkg}^{i} as "perfectly known" background
    - Distribution of signal counts as Poissonian (F_{sig}^{i}) with $\mu = n_{sig}^{i}$
    - Find n_{sig}^{i}, such that the Poissonian $\left(F_{sig+bkg}^{i}\right)$ with $\mu = n_{bkg}^{i} + n_{sig}^{i}$ has $Prob\left[F_{sig+bkg}^{i}(x) > n_{ISS}\right] = 95\%$
  - Plot the distribution of n_{sig}^{i} and take the 95% as a final result

Confidence limit distribution (AGL)



- 80% efficiency on R>0 AE anomaly score
- 75% efficiency on CL score $\left|\frac{R_{UH}-R_{LH}}{R_{UH}+R_{LH}}\right| < 0.2$