CBM performance for (multi-)strange hadron measurements using Machine Learning techniques

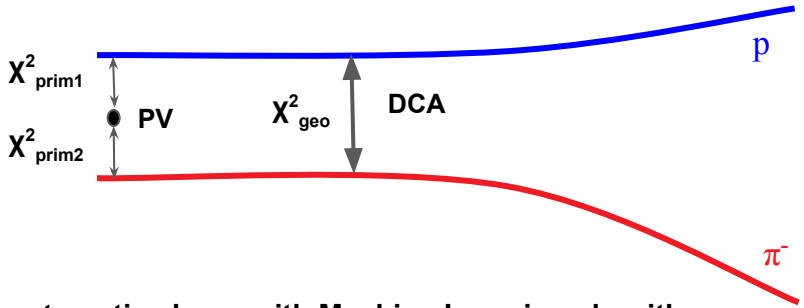
Shahid Khan, Olha Lavoryk, Oleksii Lubynets, Viktor Klochkov, Andrea Dubla, Ilya Selyuzhenkov

for the CBM Collaboration

- The production of strange quarks is sensitive to the properties of created matter in heavy-ion collisions
- CBM, due to its high interaction rate capability, has the possibility of reconstructing rare multi-strange particles and hypernuclei
- Λ hyperon is the most abundantly produced strange baryon at FAIR energies
- For CBM performance studies use, collisions generated with URQMD and DCM-QGSM-SMM:
  Au+Au collisions at p_{beam} = 12*A* GeV/*c* ($\sqrt{s_{NN}}$ = 4.93), mbias, 600k, Multiplicity bin (200-400)
- CBM simulation: GEANT4 Monte Carlo, CA tracking, KFParticle within CbmRoot framework

$\Lambda^0 \rightarrow p + \pi^-$ decay reconstruction parameters:

- Variables associated with tracks are used for the optimization of selection criteria :
  - **χ^2_{prim}, DCA, χ^2_{geo}, L/ΔL**



Selection criteria are optimized multi-dimensionally, non-linearly and in an automatized way with Machine Learning algorithms

Boosted Decision Trees (XGBoost Library) Implementation & Yield extraction

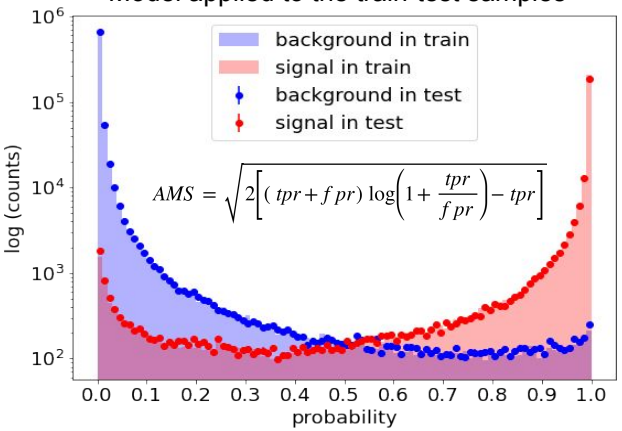
Data preparation and model deployment:

- DCM-QGSM-SMM sample as simulated data (MC signal in the 5σ around peak)
- UrQMD sample is treated as experimental data (combinatorial background outside the peak region)
- BDT model is trained and tested on separate Λ candidate samples
- Apply the XGB trained-tested model on 600k events of URQMD (yield extraction) and DCM (efficiency calculation)
  - best BDT is selected on Approximate Median Significance (AMS)

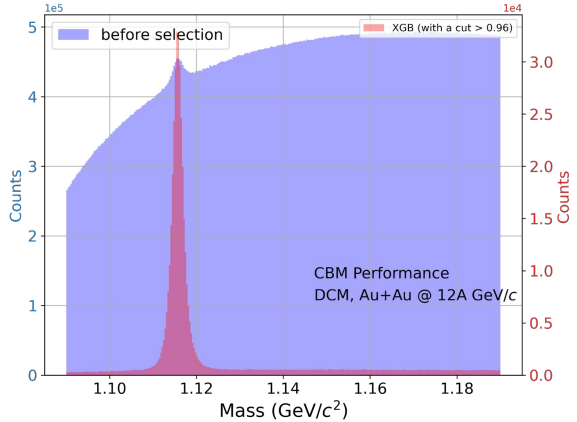
Yield Extraction Procedure:

- Signal shape: Double Sided Crystal Ball (DSCB)
- Background shape: 2nd order polynomial

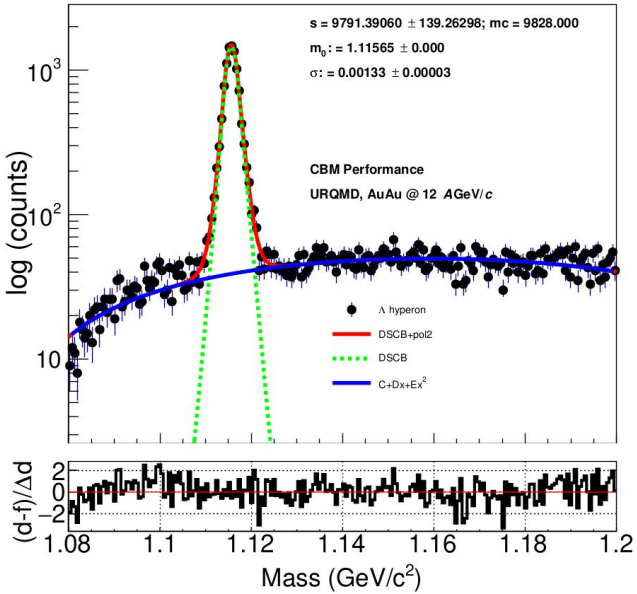
Model applied to the train-test samples



Apply the XGB model on URQMD events



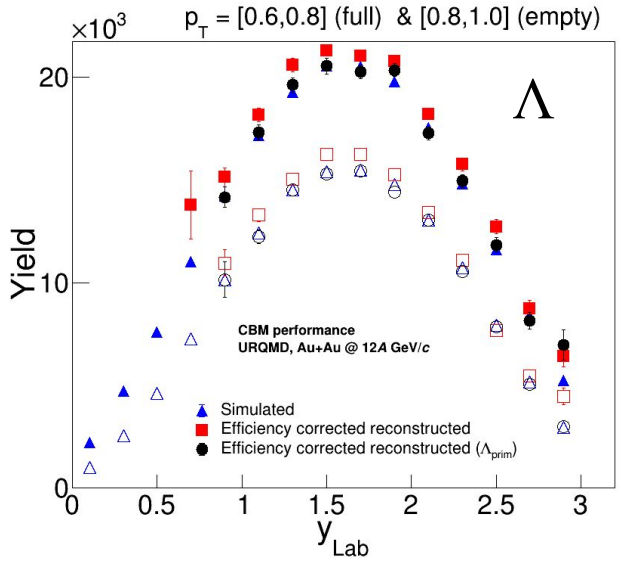
Fitting procedure



Performance of the Λ yield extraction

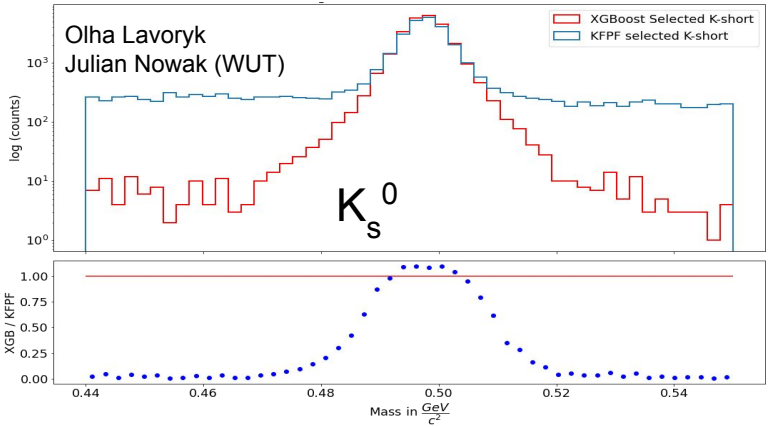
Corrected yield of primary Λ (**black circles**) reproduces simulated input (**blue triangles**)

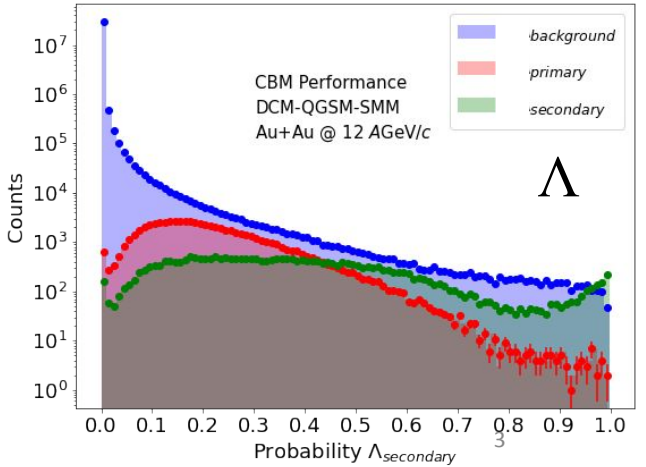
10-20% excess in the extracted Λ yield (**red squares**)→ requires feed-down correction



Outlook

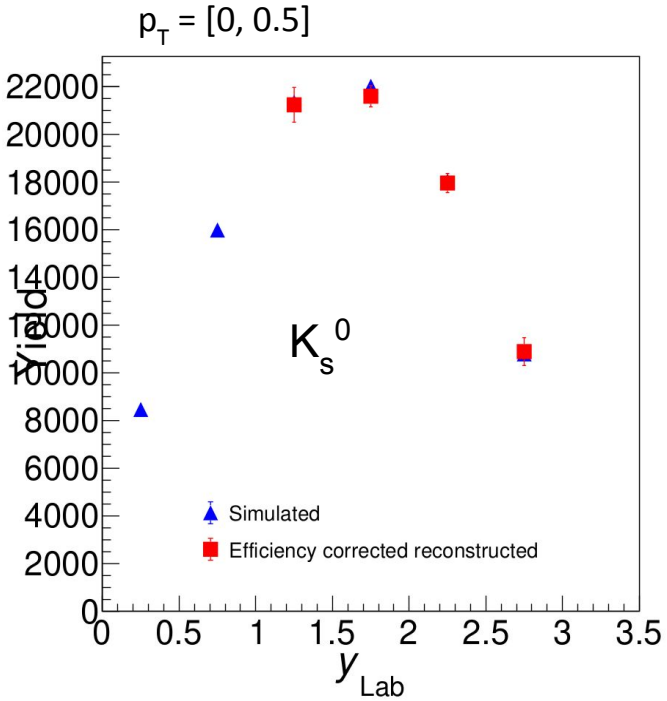
- Multi-classifier BDT to separate primary and secondary Λ
- Evaluate systematic uncertainties
  - XGB selection variation
  - Yield extraction procedure
- ML application for yield measurement for $K_s^0 \rightarrow \pi^+\pi^-$

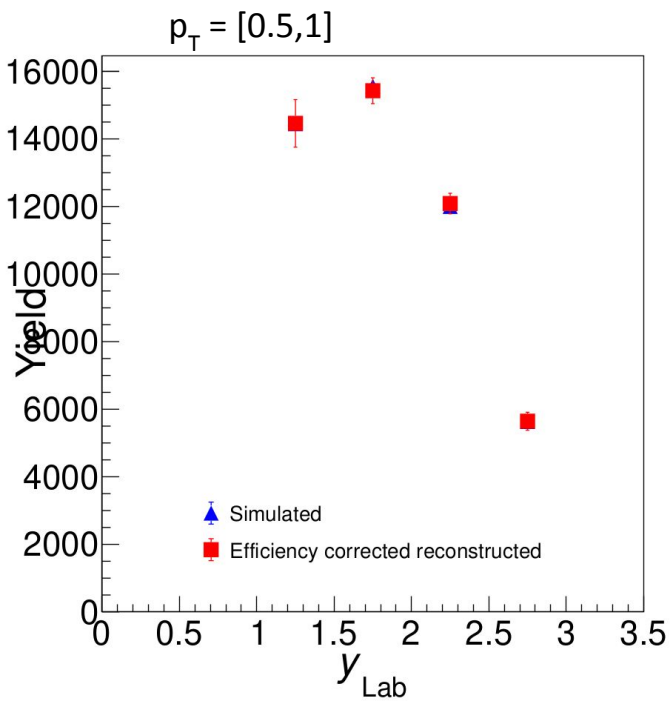




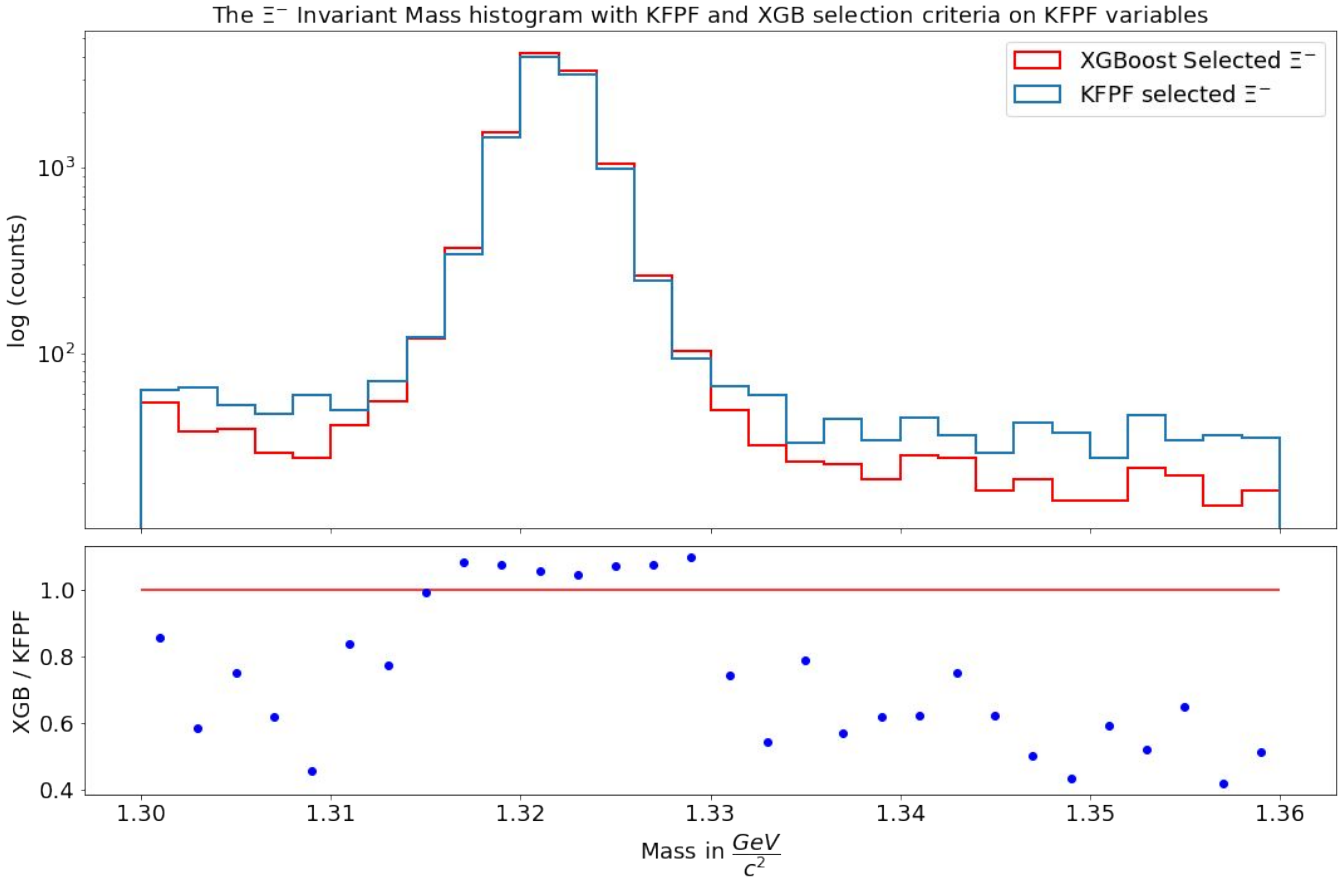
Back up

SQM Proceedings

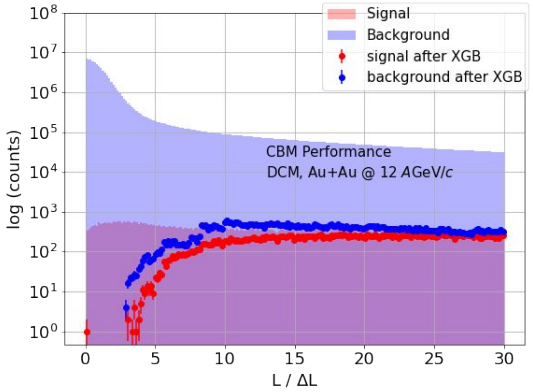

p_T = [0, 0.5]
Yield
K_s^0
Simulated
Efficiency corrected reconstructed
y_{Lab}

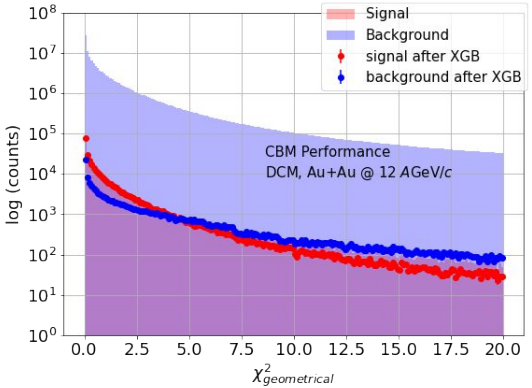

p_T = [0.5,1]
Yield
Simulated
Efficiency corrected reconstructed
y_{Lab}

Ξ⁻ Reconstruction

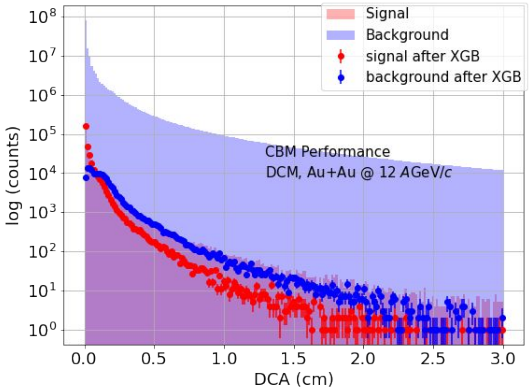


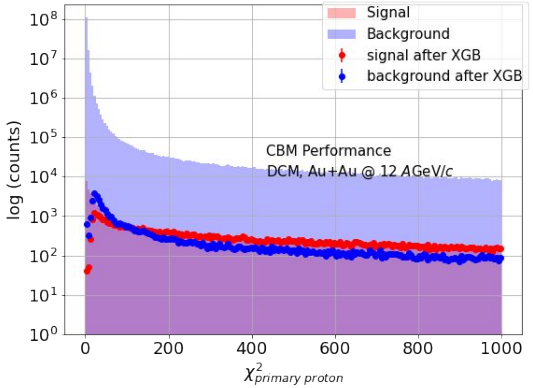
Distribution of True signal and background in XGB Selected Signal

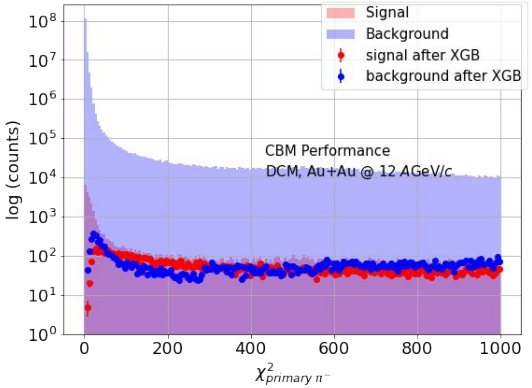




Distribution of MC signal (pure signal) and background in the XGB selected signal

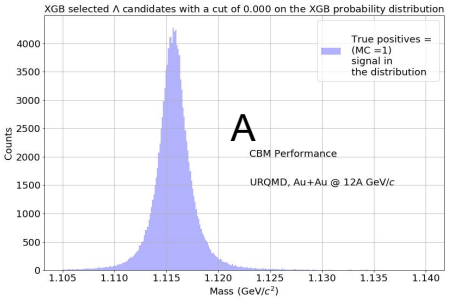


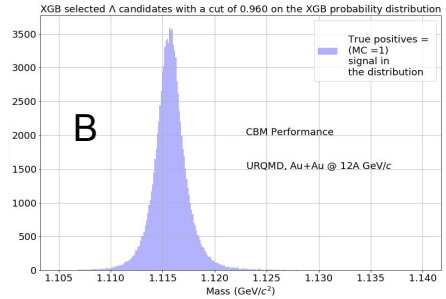


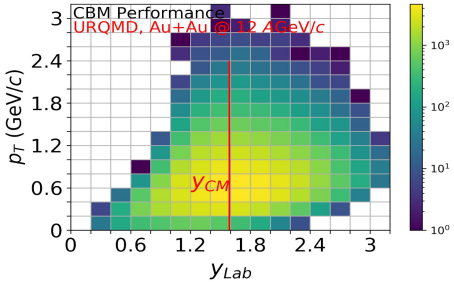


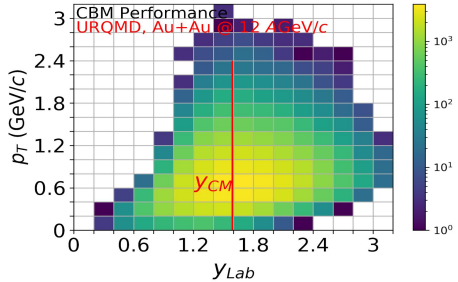
XGB performance for Λ candidates selection

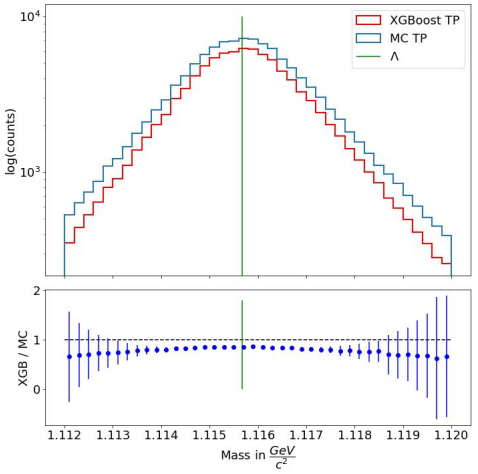
Input for efficiency (p_T, y) calculation



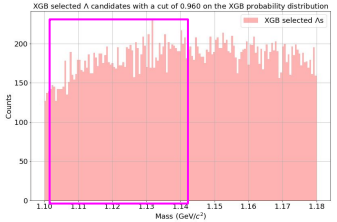


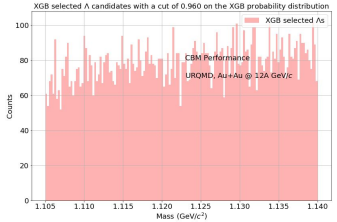






- Give the model pure background and see its predictions



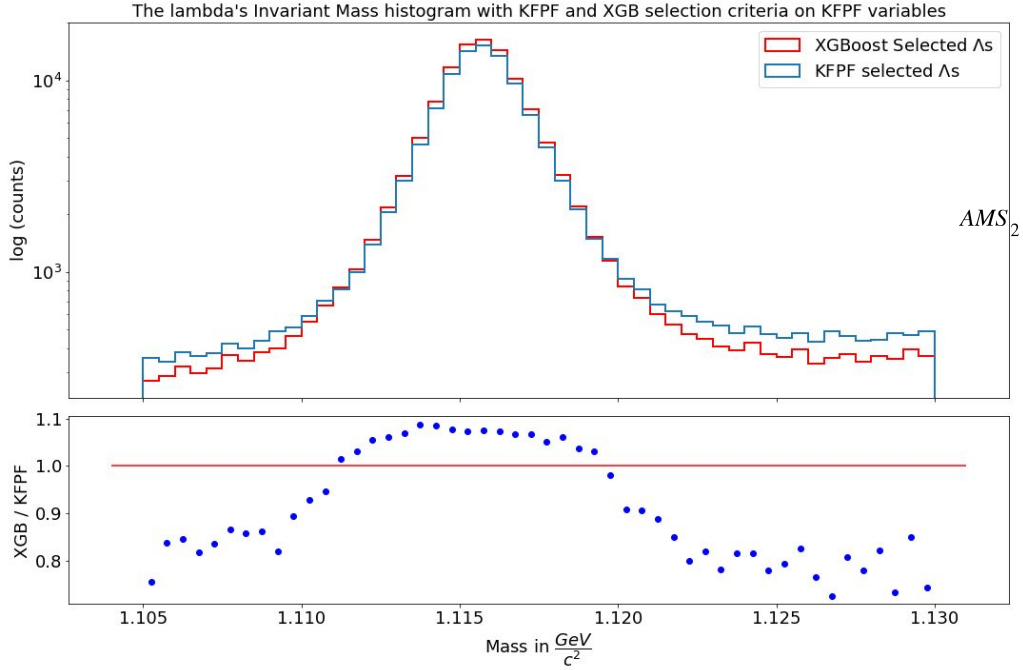


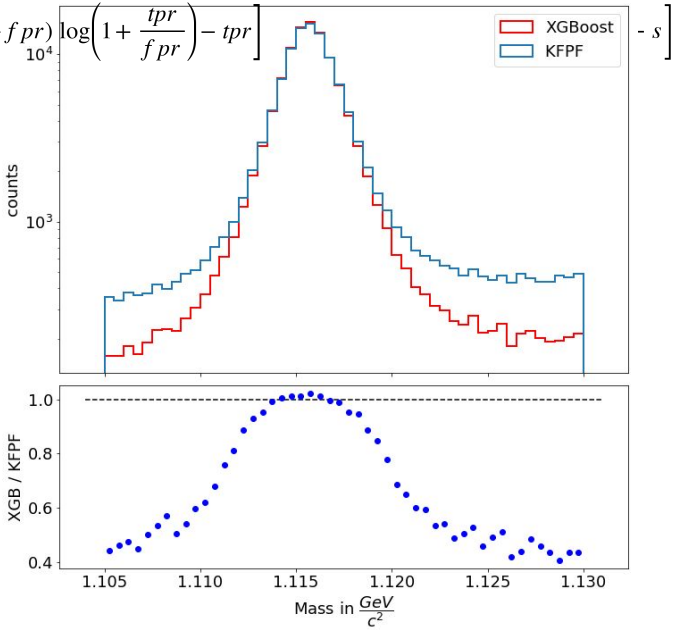
Comparison with manual selection criteria

$$AMS = \sqrt{2\left[(tpr + fpr)\log\left(1 + \frac{tpr}{fpr}\right) - tpr\right]} \quad \text{- } s\Big]$$

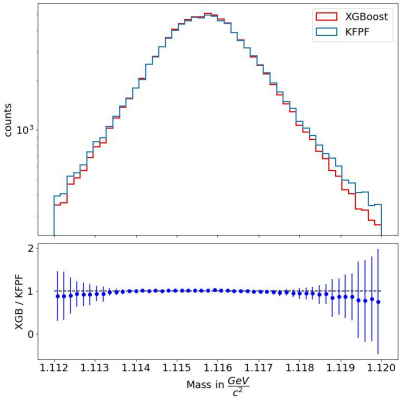
$$AMS_2 = .$$

$$AMS_2 = \sqrt{2\left[b\left\{\left(\frac{s}{b}\right.\right.\right.}$$





$$AMS_2 = \sqrt{2\Big[$$



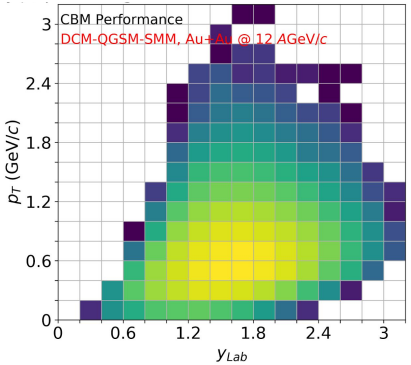
$$AMS_2 = \frac{s}{\sqrt{b}} \times \sqrt{1 + O\left[\left(\frac{s}{b}\right)^3\right]}$$

Yield Extraction

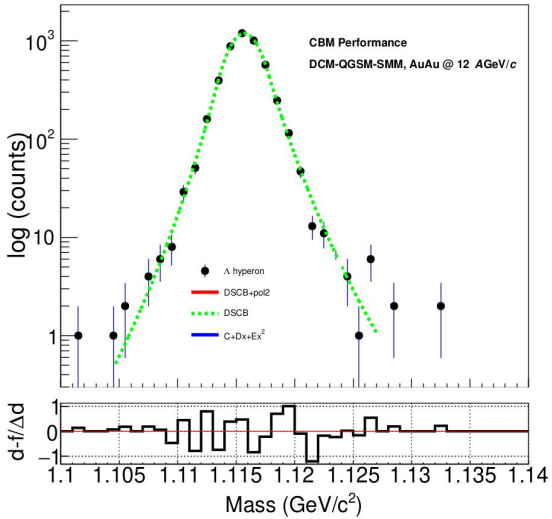
Double Sided Crystal Ball (DSCB), Gaussian with power law tails, function is used for signal and 2nd order polynomial for background approximation:

1. Fit DSCB on MC data in 4σ around the mean
2. Exclude signal region (m<1.108 & m>1.13) and fit background with *pol2(m)*
3. Use DSCB+*pol2* as a fit function on the total range of data and initialize the fit parameters by step 1 and 2

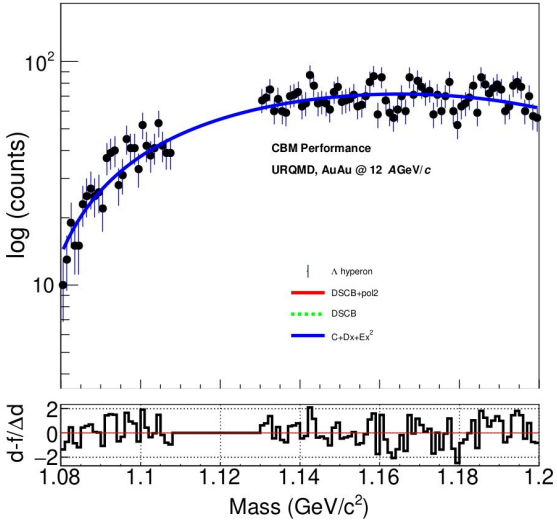
Divide (p_T,y) phase space into 15x15 bins



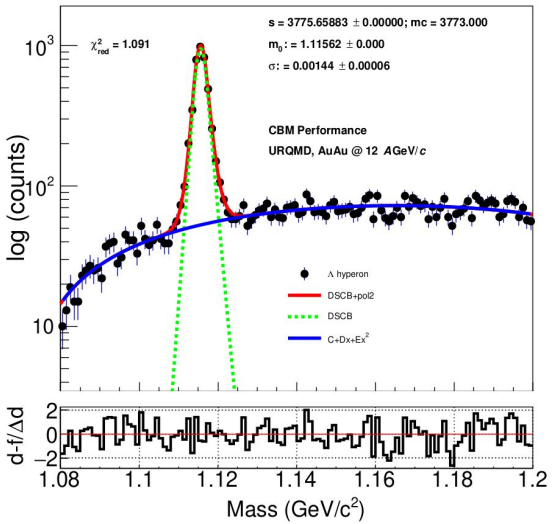
Step 1



Step 2



Step 3



DSCB

Parameters of the fitting function

μ & σ parameters of the gaussian

a_1 & n_1 parameters of the left power law tail

a_2 & n_2 parameters of the right power law tail

A_0 Co-efficient

Variables inside the fitting function

$$u = (x - 1.1157 - \mu)/\sigma$$

$$A_1 = (n_1 a_1)^{n1} \times e^{-a_1^2/2}$$

$$A_2 = (n_2 a_2)^{n2} \times e^{-a_2^2/2}$$

$$B_1 = n_1/a_1 - a_1 \qquad B_2 = n_2/a_2 - a_2$$

Conditions

If $u < -a_1$ then $A_0 \times A_1(B_1 - u)^{-n_1}$

If $u \geq -a_1$ & $u<a_2$ then $A_0 \times e^{-u^2/2}$

If $u \geq a_2$ then $A_0 \times A_2(B_2 - u)^{-n_2}$

Parameters' bounds

a_1 [0,10] ; n_1[0,100]; a_2 [0,10] ; n_2[0,100];