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

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

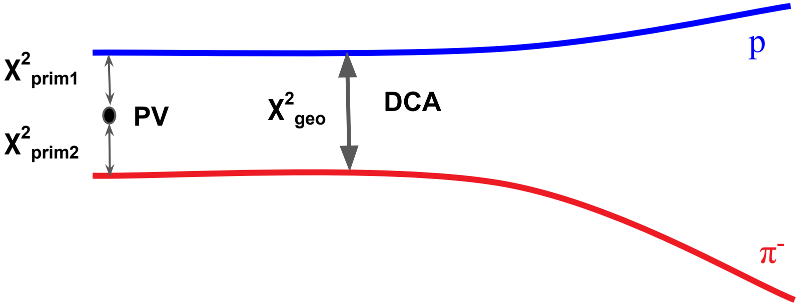
for the CBM Collaboration

- 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:

- χ^2_{prim} - squared distance between the daughter track and the primary vertex divided by its Covariance Matrix (CV)
- **DCA** - distance of closest approach between proton & pion tracks
- χ^2_{geo} - squared distance between daughter tracks divided by CV
- **L/ΔL** - distance between primary and secondary vertex divided by CV



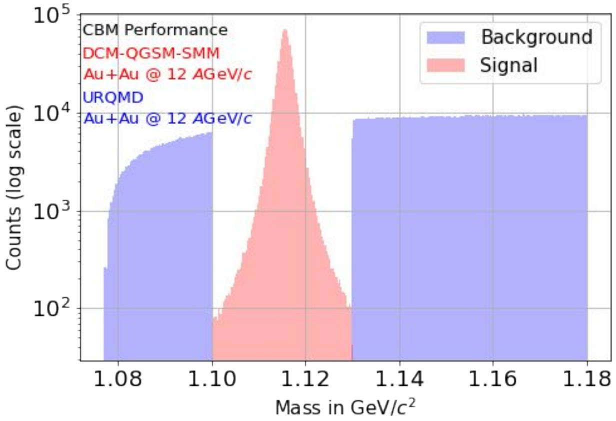
Selection criteria are optimized multi-dimensionally, non-linearly and in an automatized way with Machine Learning algorithms

Boosted Decision Trees (XGBoost Library) Implementation

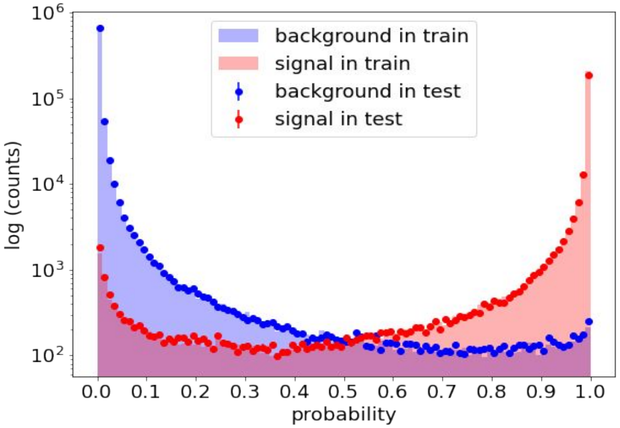
Data preparation:

- DCM-QGSM-SMM sample as simulated data (MC signal)
- UrQMD sample is treated as experimental data (MC background)
- Λ candidates sample is cleaned by removing those with non physical values
- Λ candidates are divided into train and test samples
- BDT model is trained on train set and then applied on train and test sets, separately

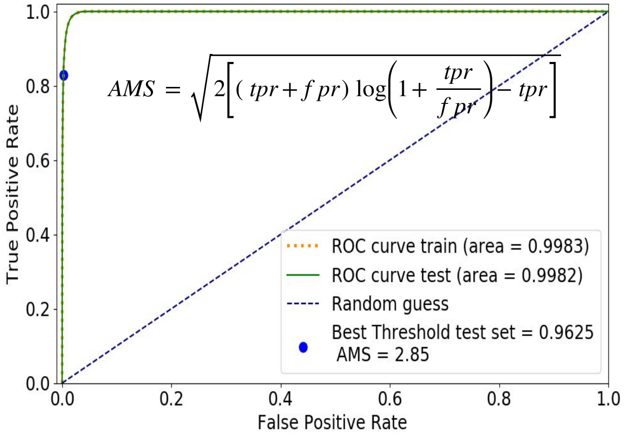
Train-test dataset



Model applied to the train-test samples

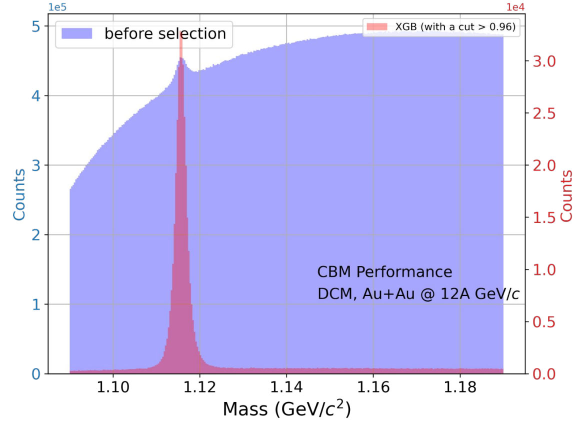


Optimize Λ candidates selection for significance via Approximate Median Significance (AMS)

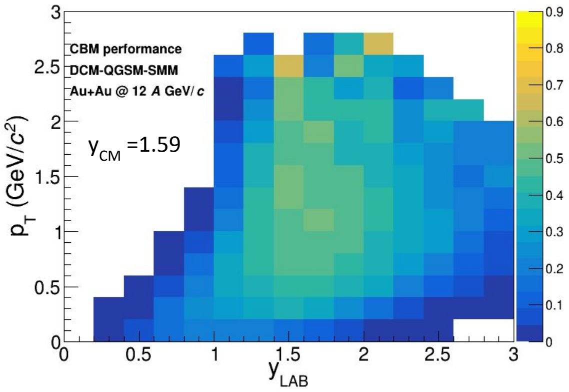


Acceptance x Efficiency of the Λ reconstruction & selection

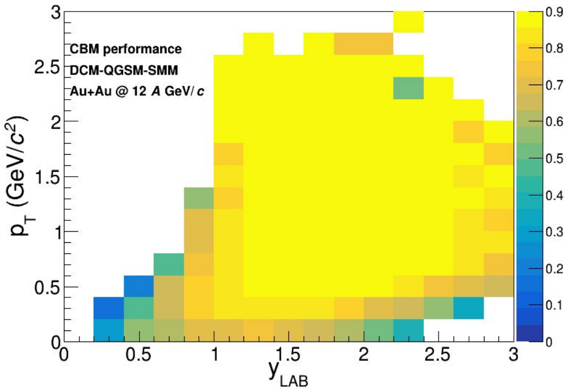
- Apply the XGB trained-tested model on 600k events of URQMD and DCM
- Calculate the reconstruction efficiency (acceptance x preselection x ML efficiency) on DCM model by dividing reconstructed yield by simulated yield



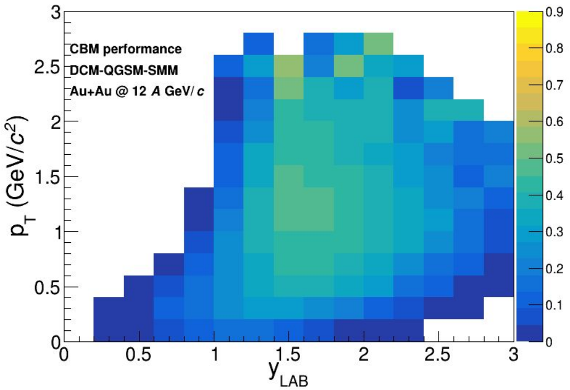
acceptance x preselection efficiency



ML efficiency



acceptance x preselection efficiency x ML efficiency



Yield extraction procedure

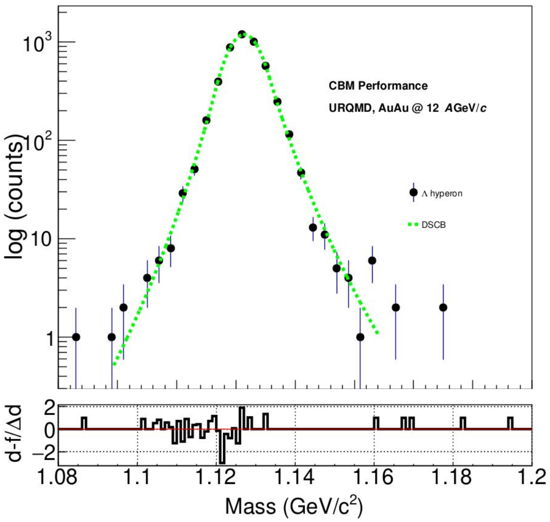
Signal shape: Double Sided Crystal Ball (DSCB) function: Gaussian with power law tails

Background shape: 2nd order polynomial

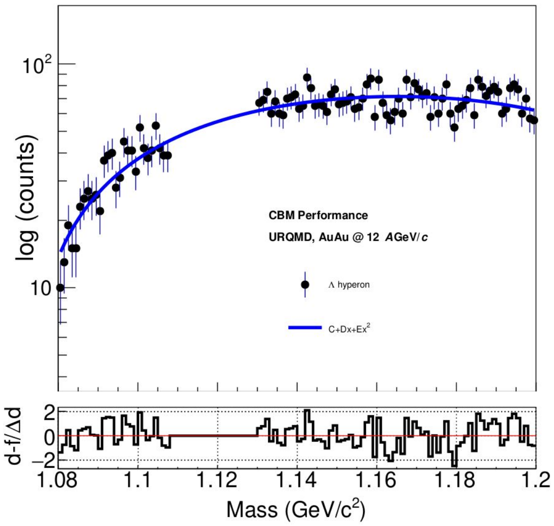
Fitting procedure

1. Fit DSCB to the MC signal distribution within 4σ around the mean
2. Fit background with *pol2(m)* in the excluded signal region (m<1.108 & m>1.13)
3. Fit with DSCB+*pol2* within the full range of inv. mass with the fit parameters initialized by Steps 1 and 2

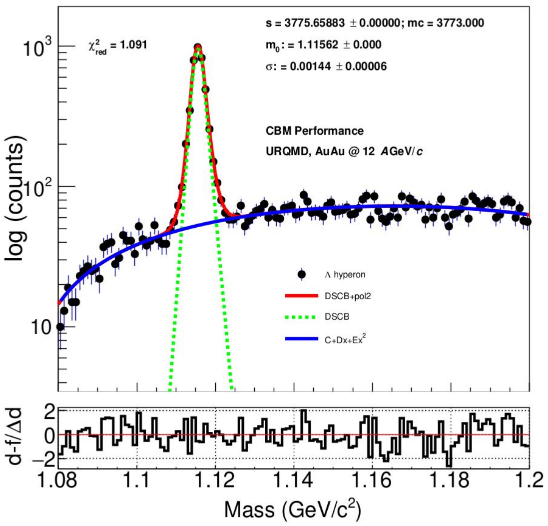
Step 1



Step 2



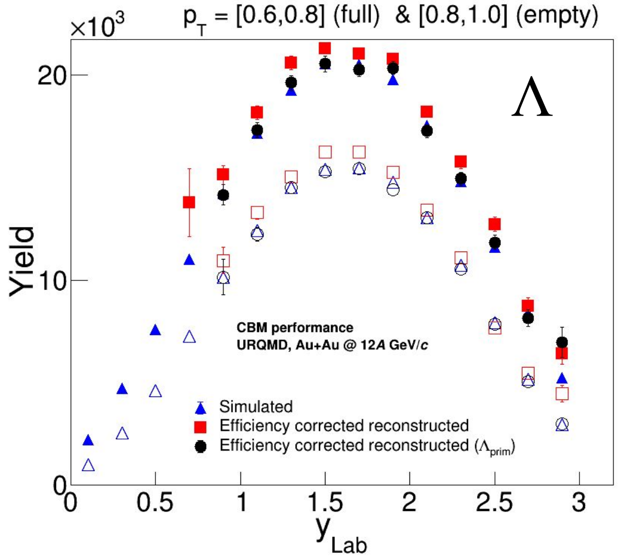
Step 3



Performance of the Λ yield extraction

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

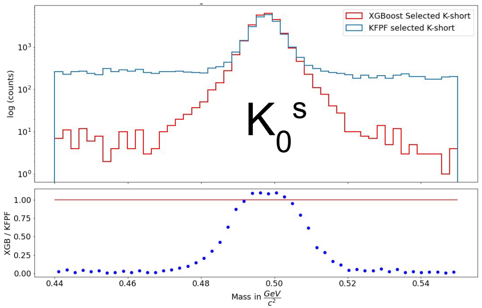
10-20% excess in the extracted Λ yield (**red squares**)→ requires feed-down correction



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 $K_0^s \rightarrow \pi^+\pi^-$ and $\Xi^- \rightarrow \Lambda\pi^-$

Julian Nowak (WUT)



Małgorzata Karabowicz (WUT)