Machine learning pipeline for $B^0_{(s,d)} \to D^{\pm(*)}_{(s)} K^{\pm(*)}$ decay

—

Author: Michał Kazanecki (Cracow University of Technology)
Supervisors: Wojciech Krupa, Agnieszka Obłąkowska-Mucha, Tomasz Szumlak

Outline

- Motivation
- Tools
- Signal and background proxy
- Training variables
- Classifier response and metrics
- Plots for reference channel
- Summary

Motivation

Measurement of γ angle of CKM matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}|e^{-i\gamma} \\ -|V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}|e^{-i\beta} & -|V_{ts}|e^{i\beta_s} & |V_{tb}| \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$
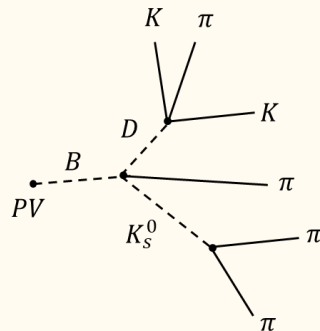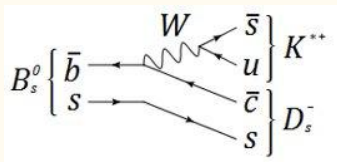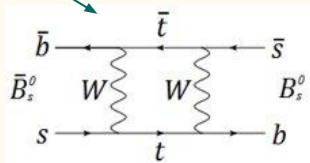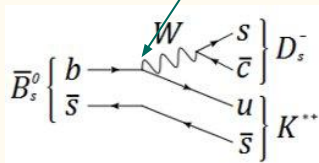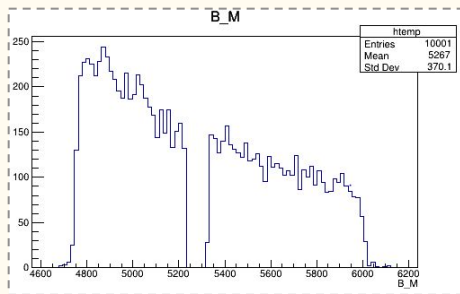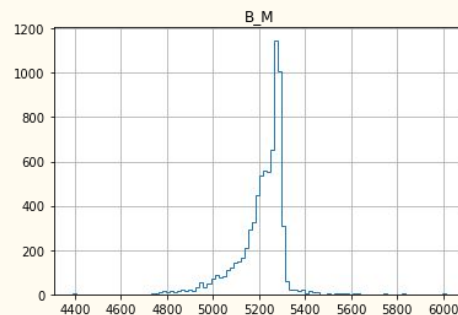
Goal: Create ML pipeline

Tools

Signal and background proxy

Background proxy - sidebands from data
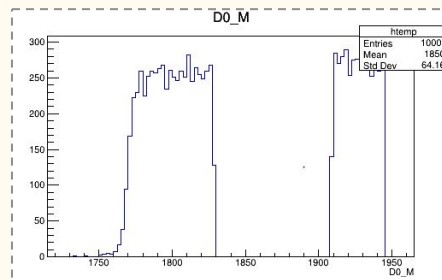
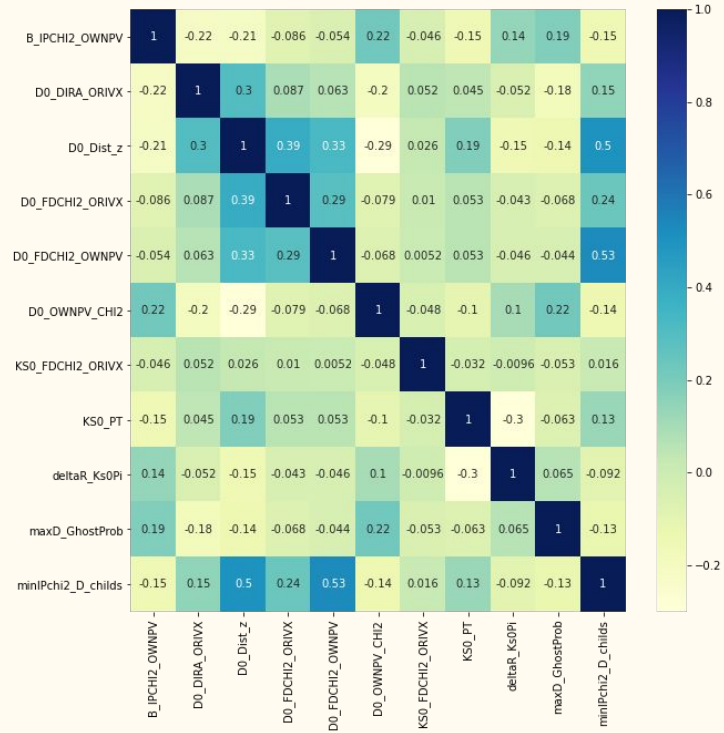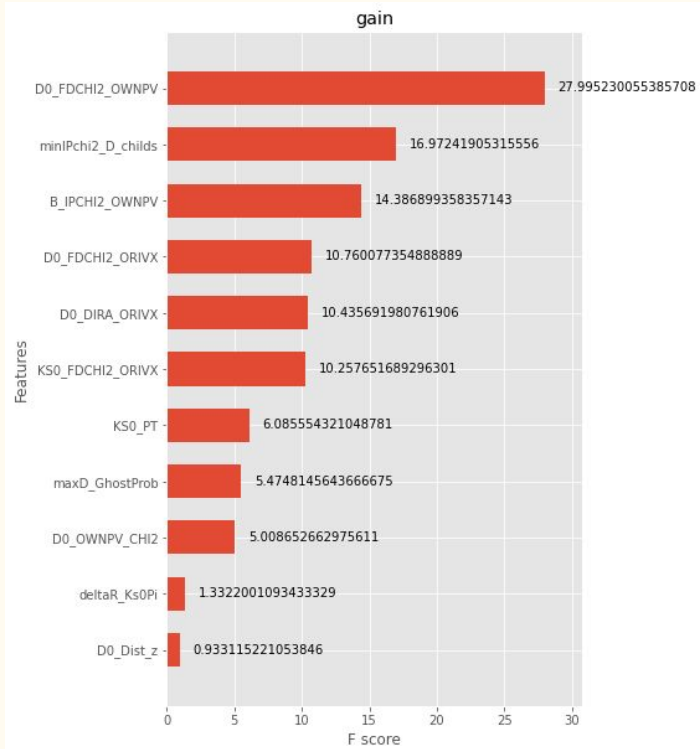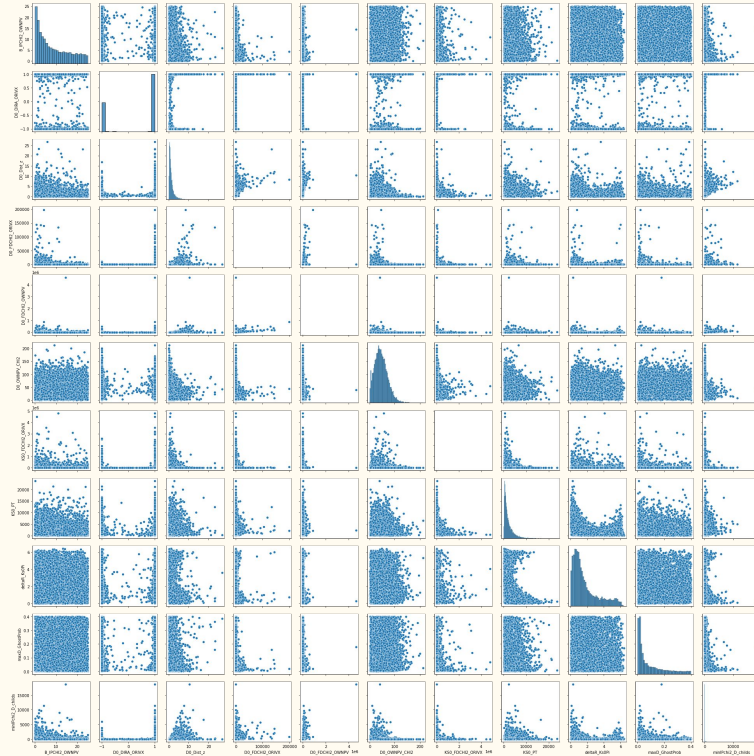
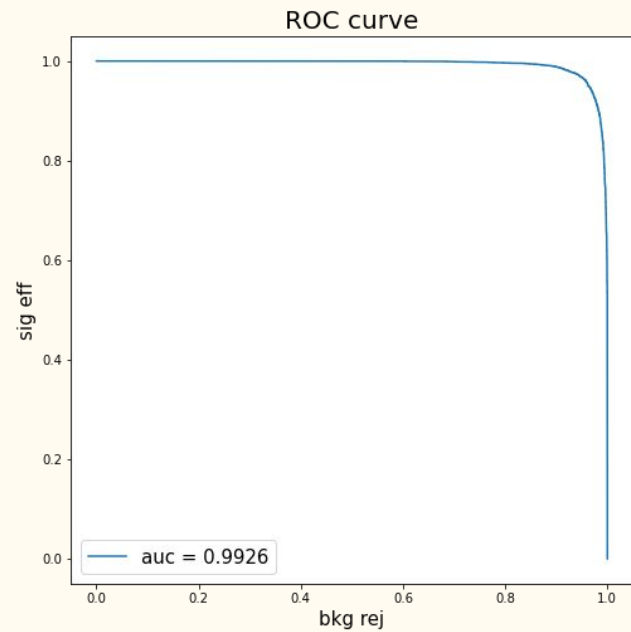Signal proxy - Monte Carlo sample

Train/test split - 50/50
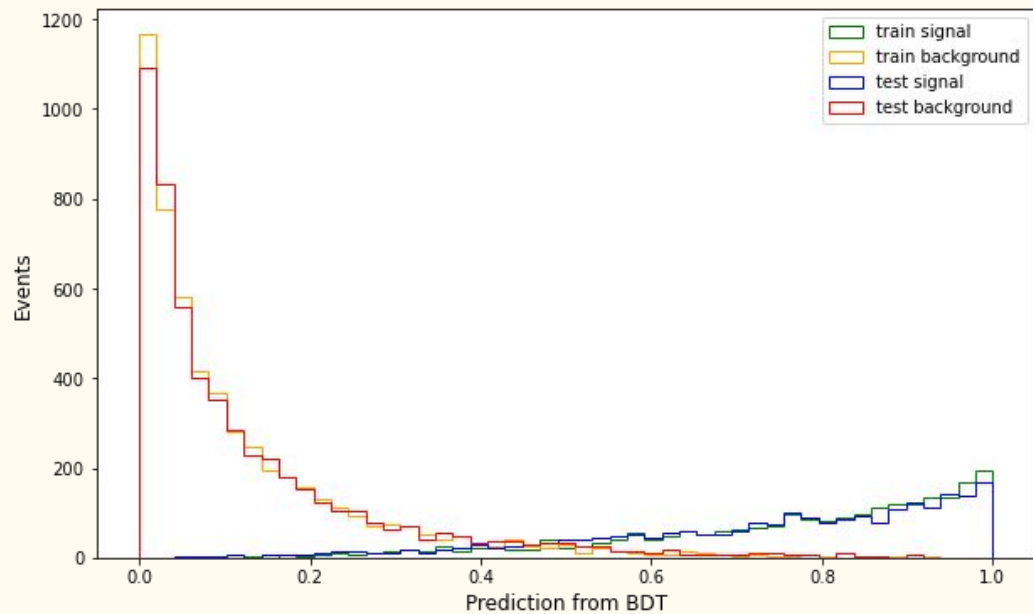
Training variables

Training variables

Classifier response and metrics

Classifier response and metrics
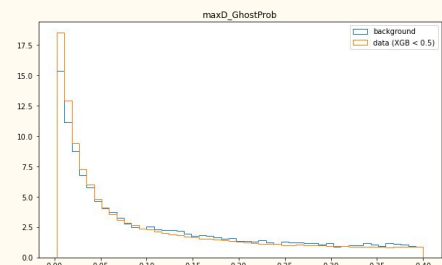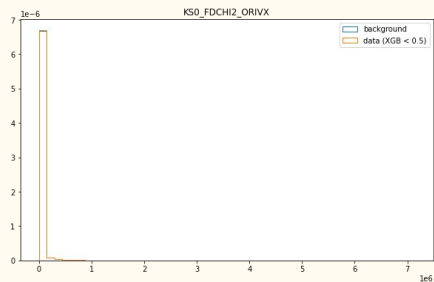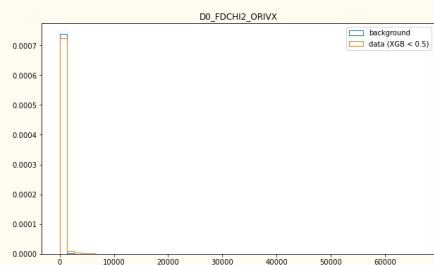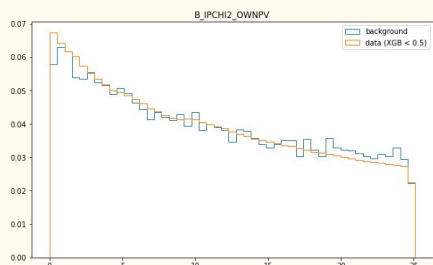
Accuracy $= 0.9612$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Precision $= 0.9423$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$$

Recall $= 0.9141$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$

F1 $= 0.9280$

$$F_1 = 2 \times \frac{\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

confusion matrix

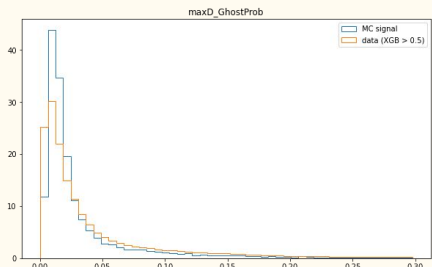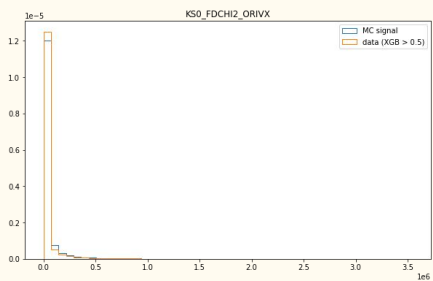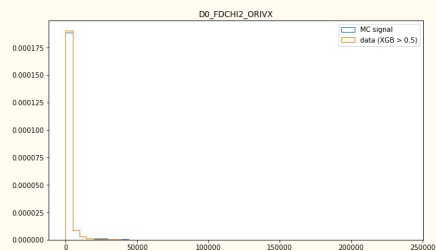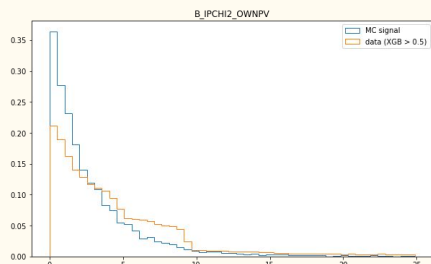|	0	1
0	True Neg 9815 71.13%	False Pos 211 1.53%
1	False Neg 324 2.35%	True Pos 3448 24.99%
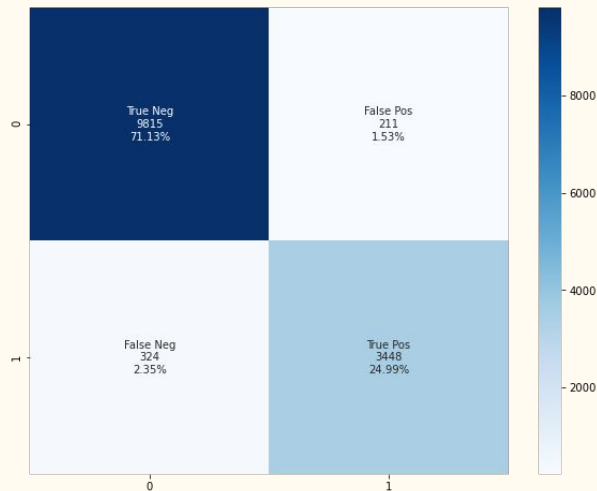
Classifier response and metrics
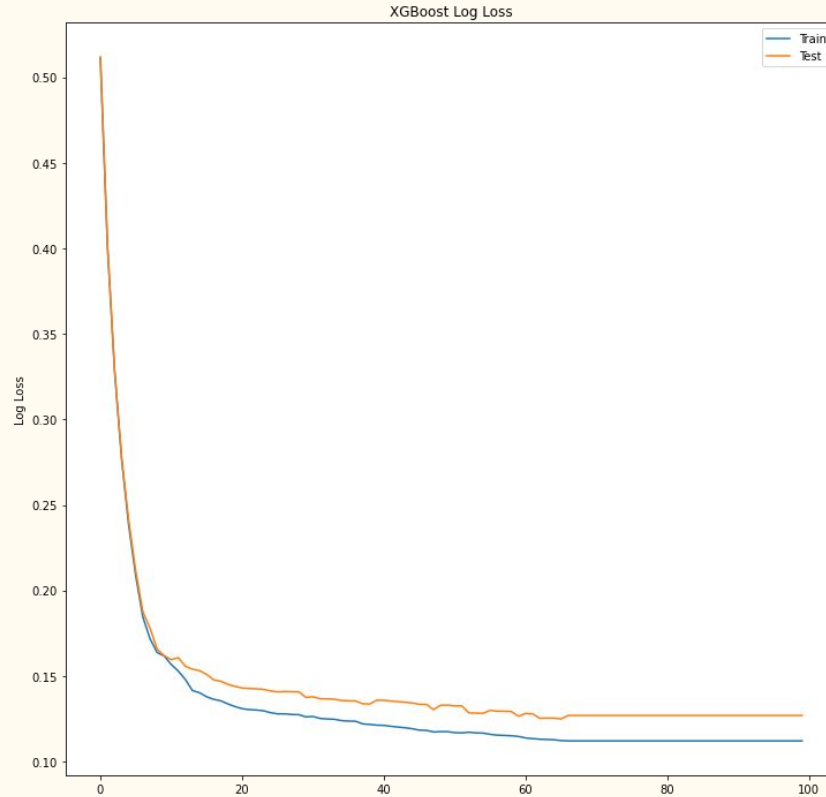
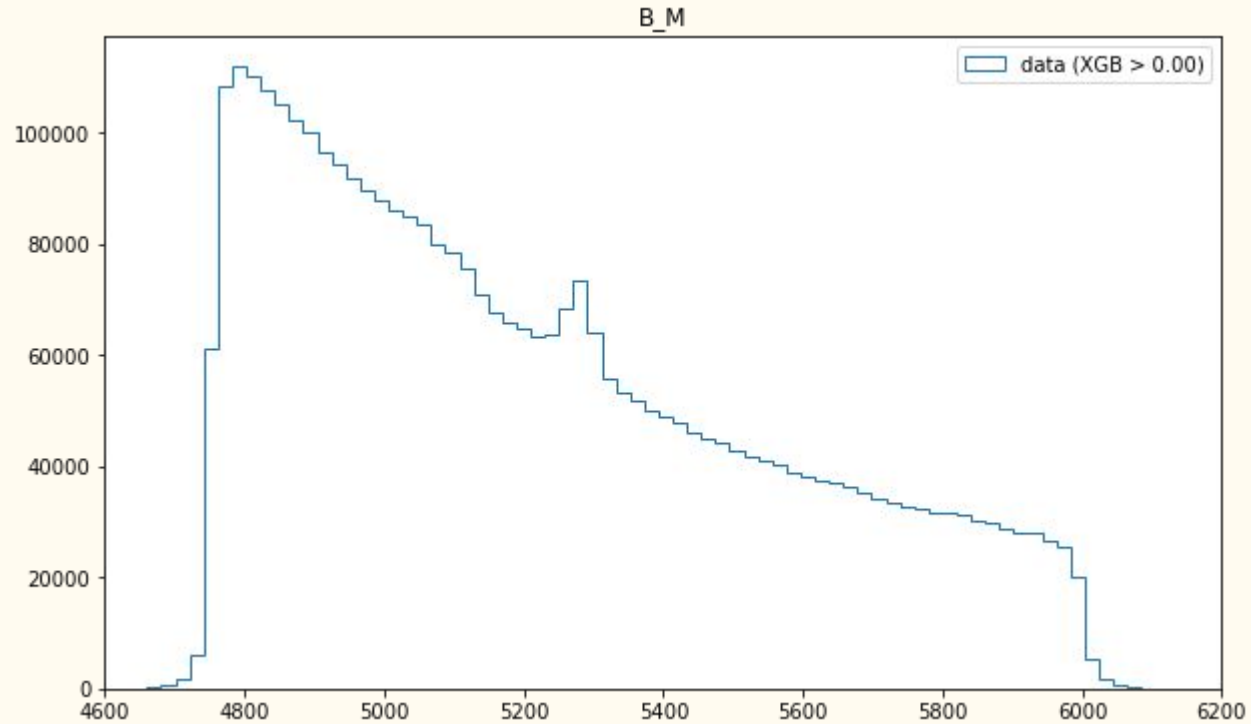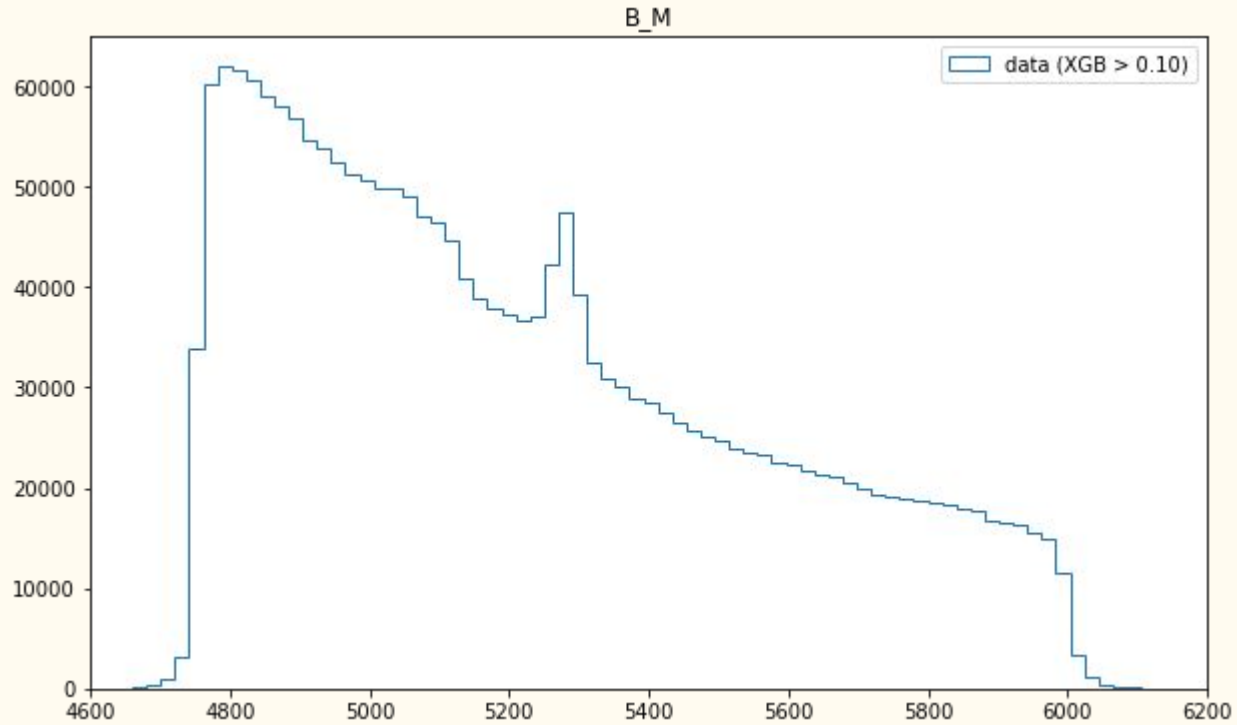Training and testing curves are very close to each other. Model is not overtrained.

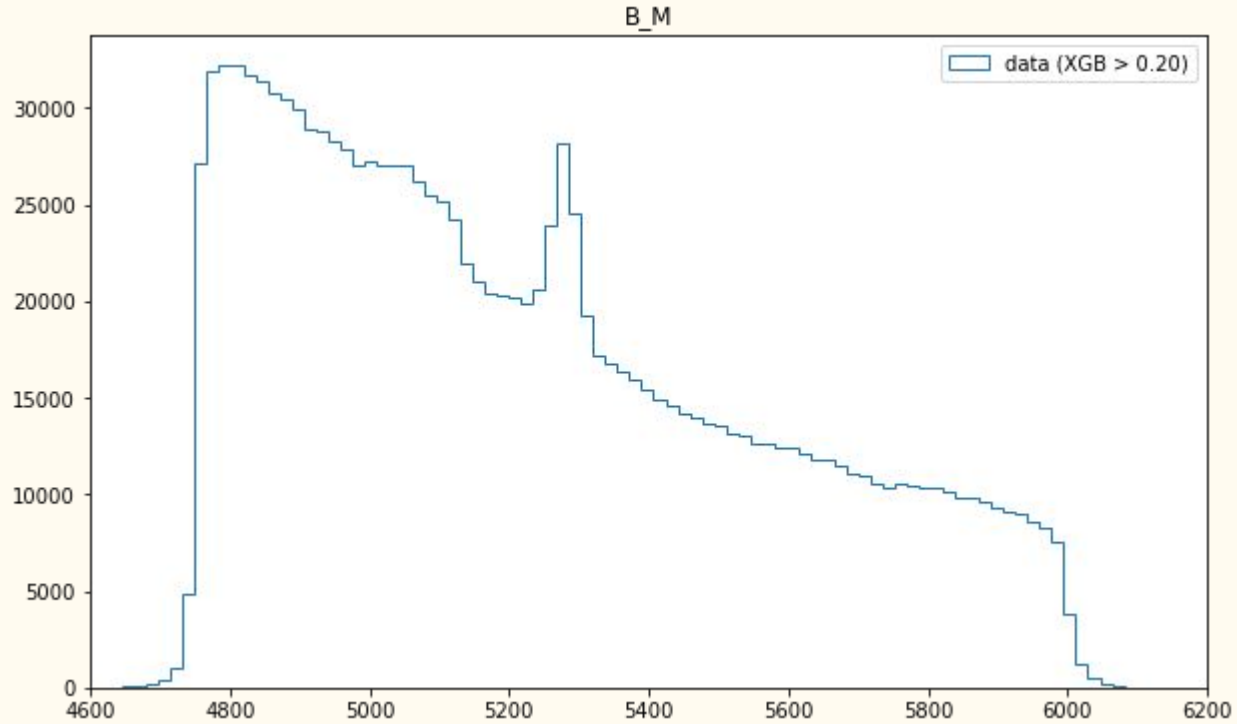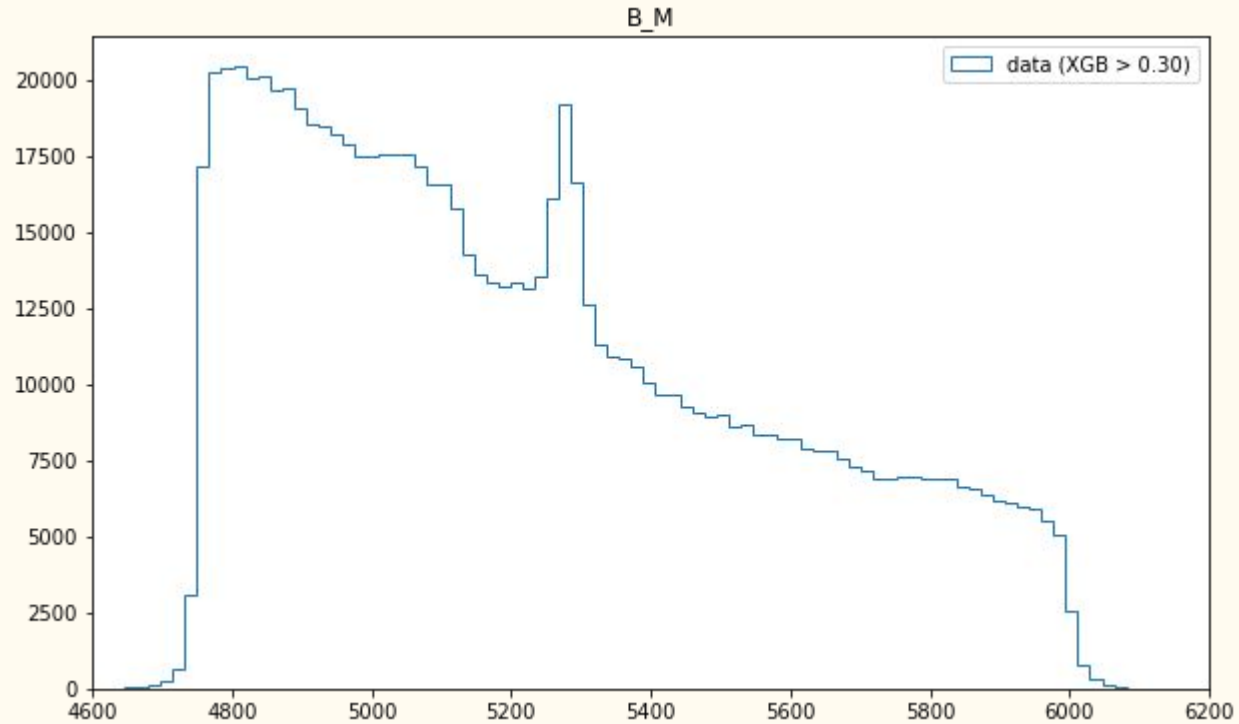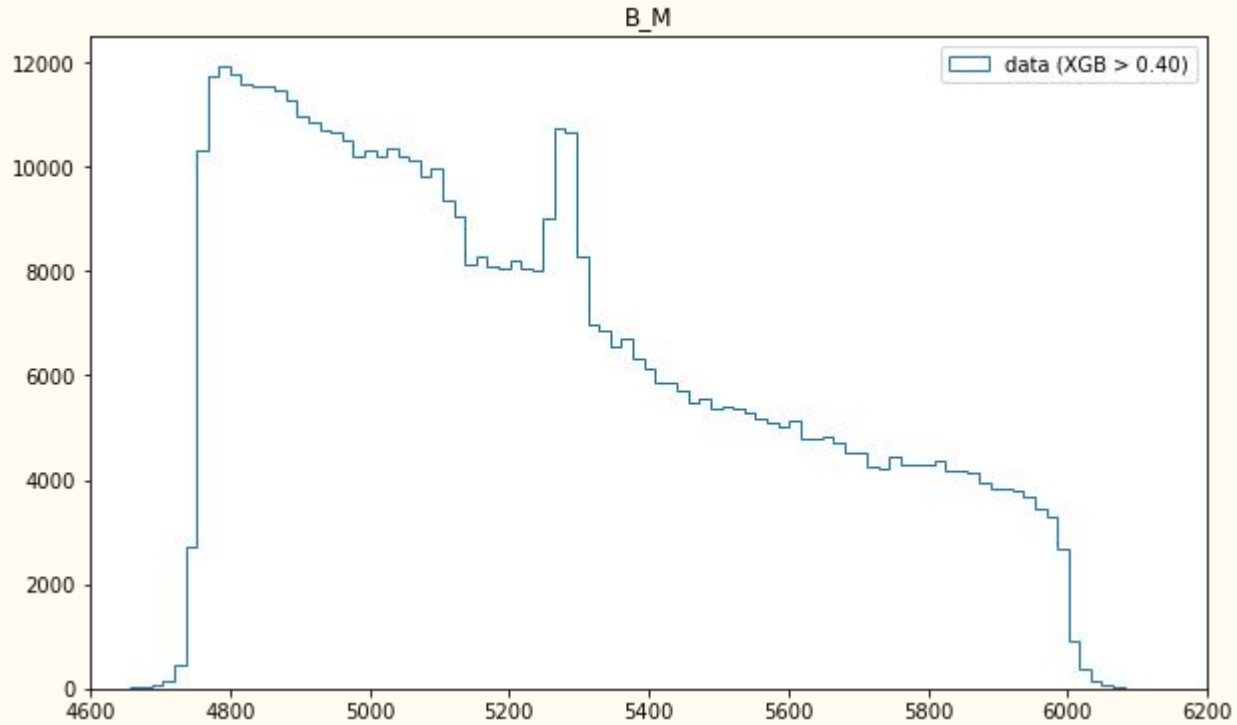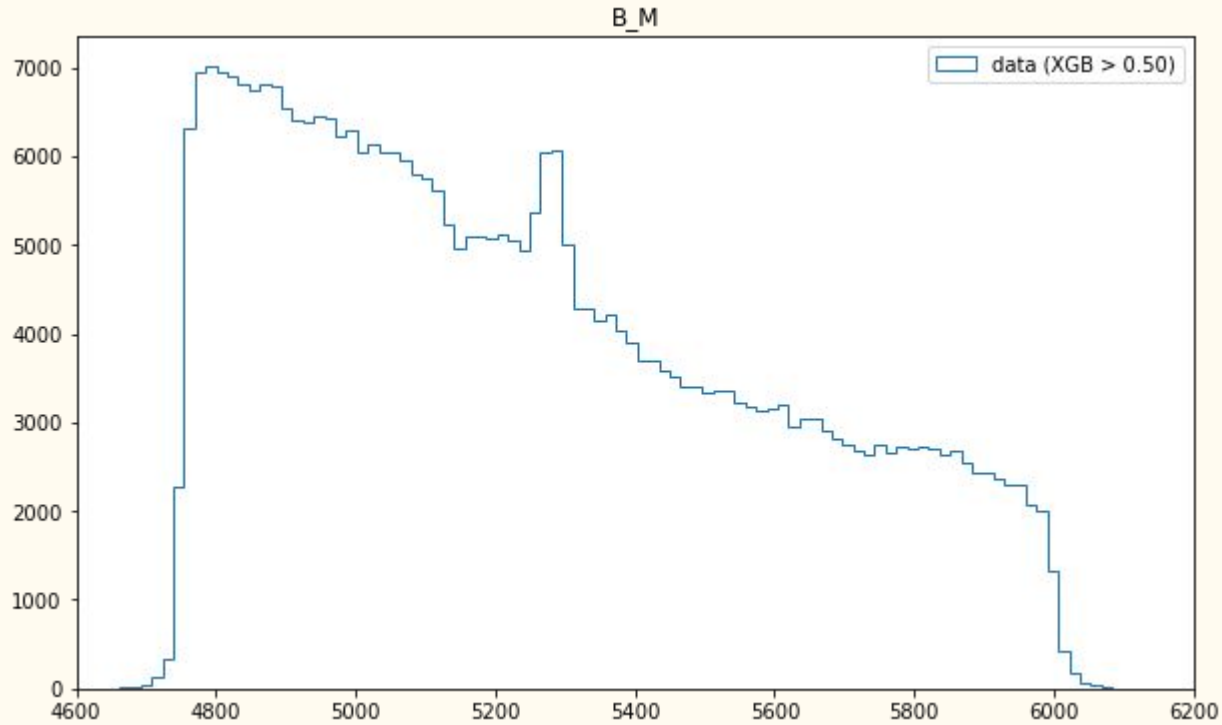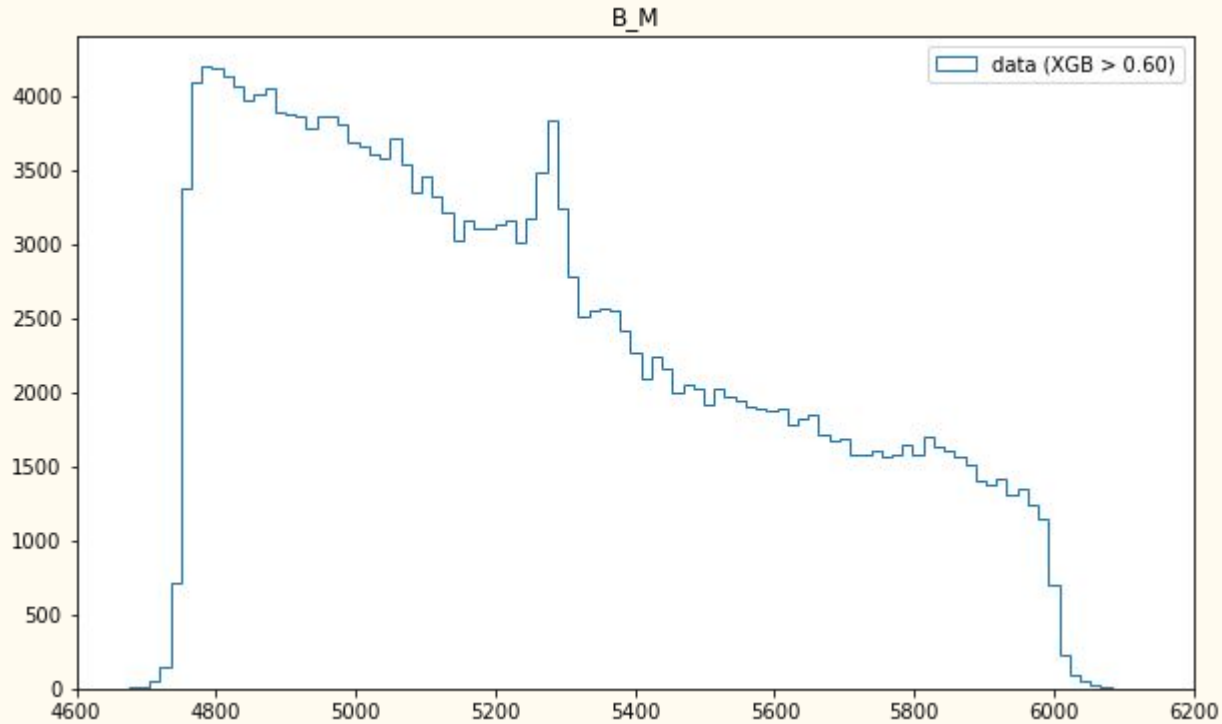# Plots for reference channel

Plots for reference channel

Plots for reference channel

B_M

Plots for reference channel

Plots for reference channel

Plots for reference channel
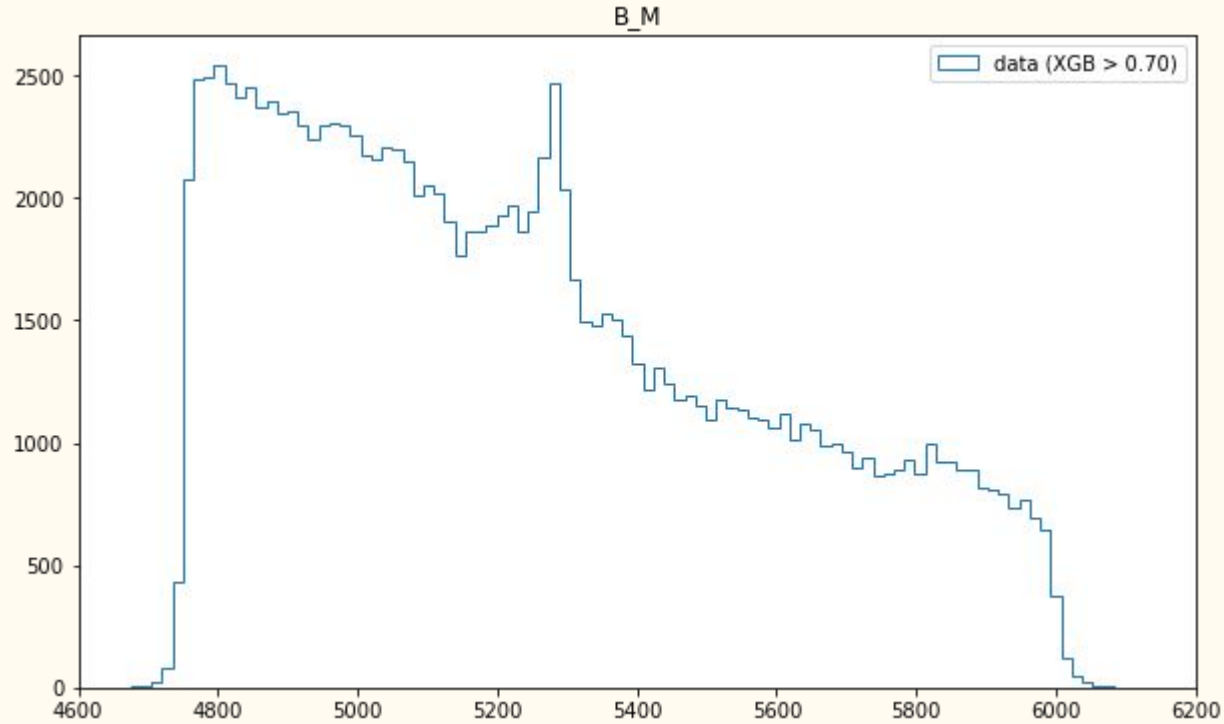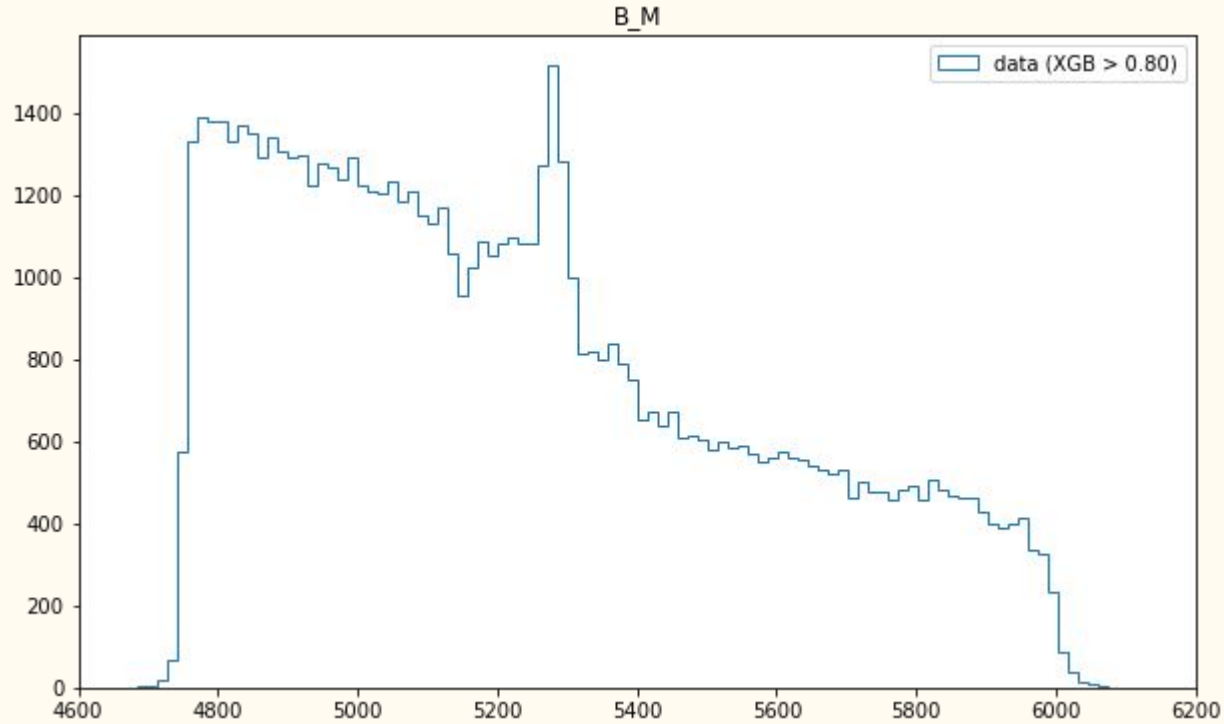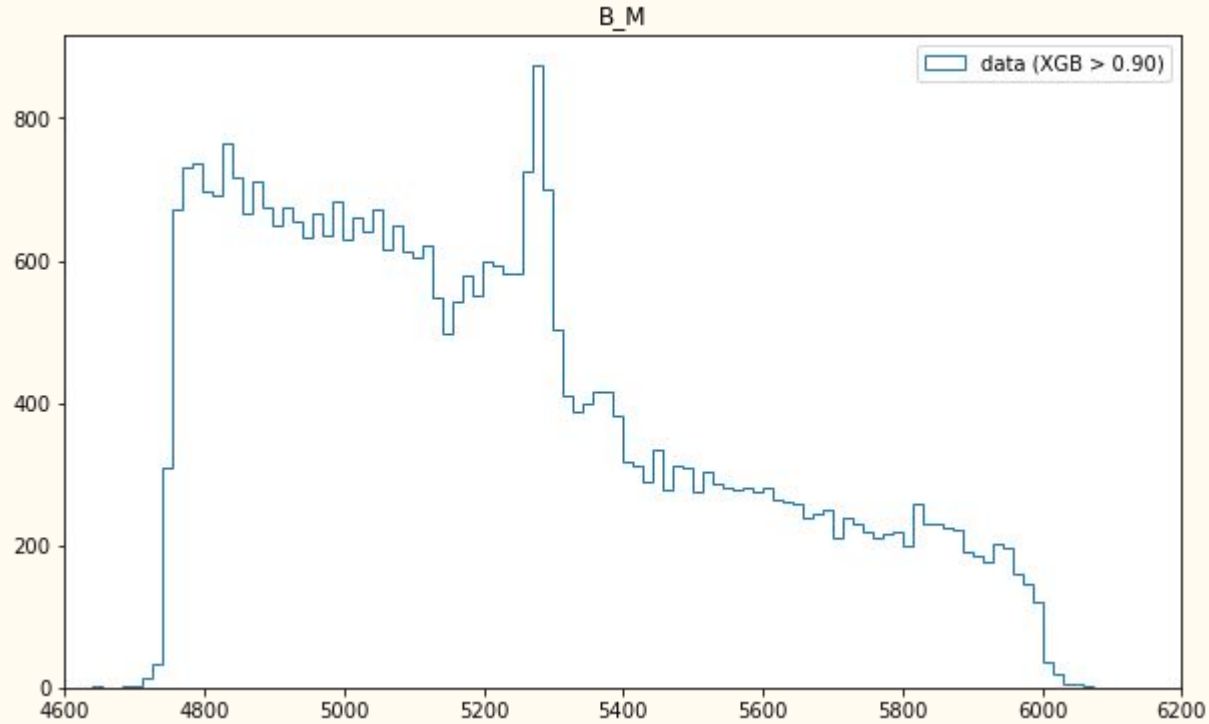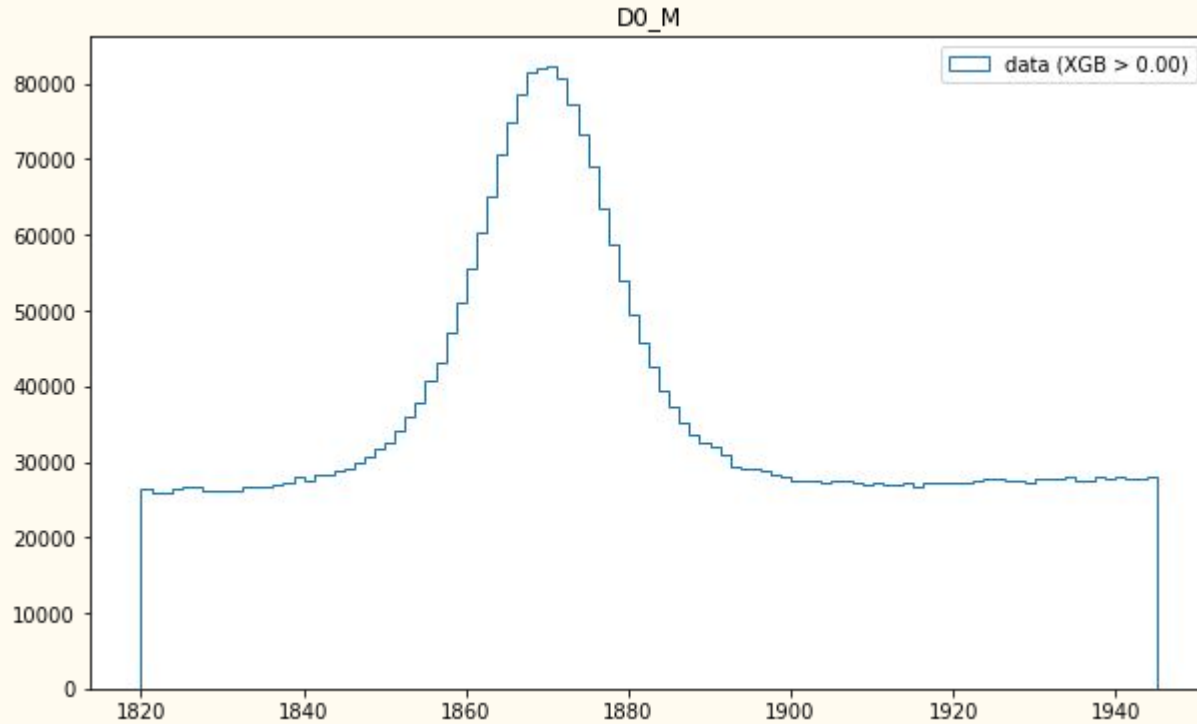

B_M

Plots for reference channel

Plots for reference channel

Plots for reference channel

Plots for reference channel

Plots for reference channel

Plots for reference channel
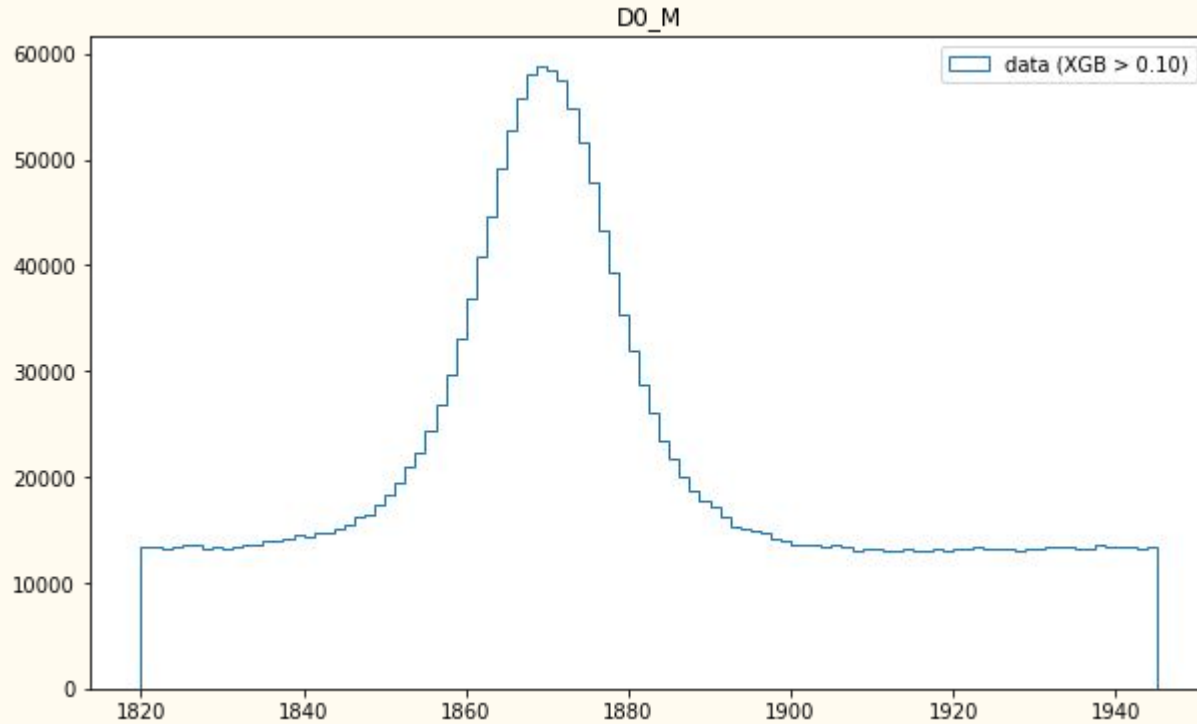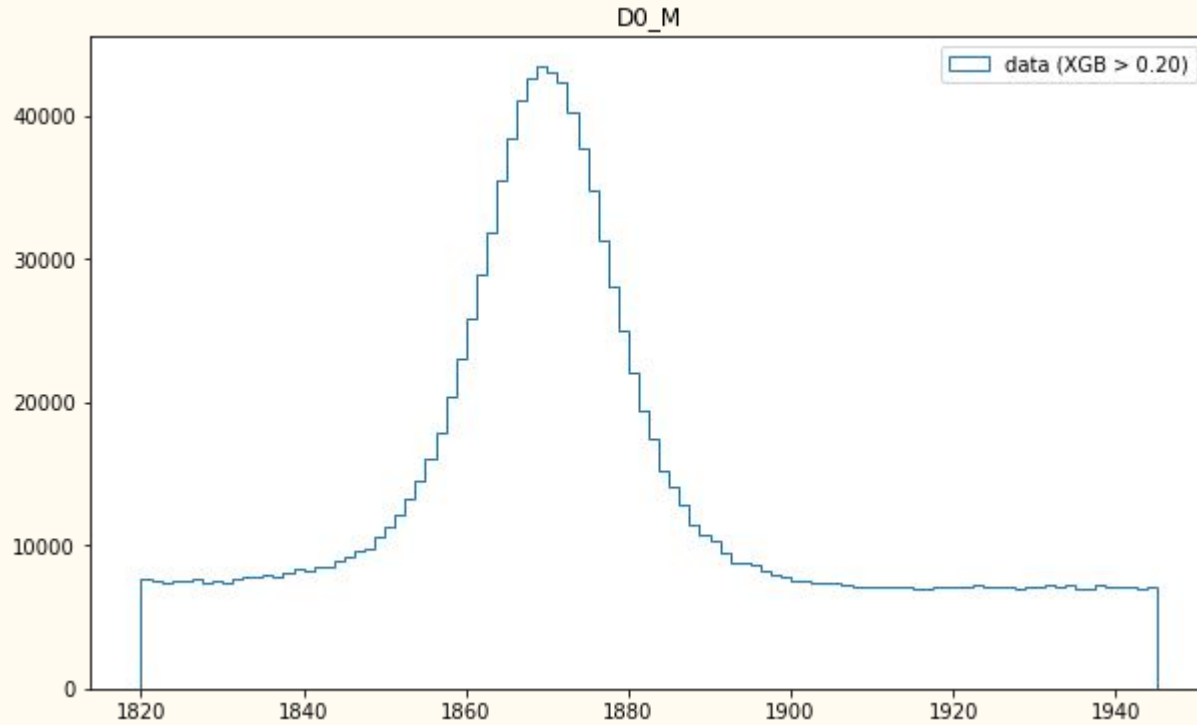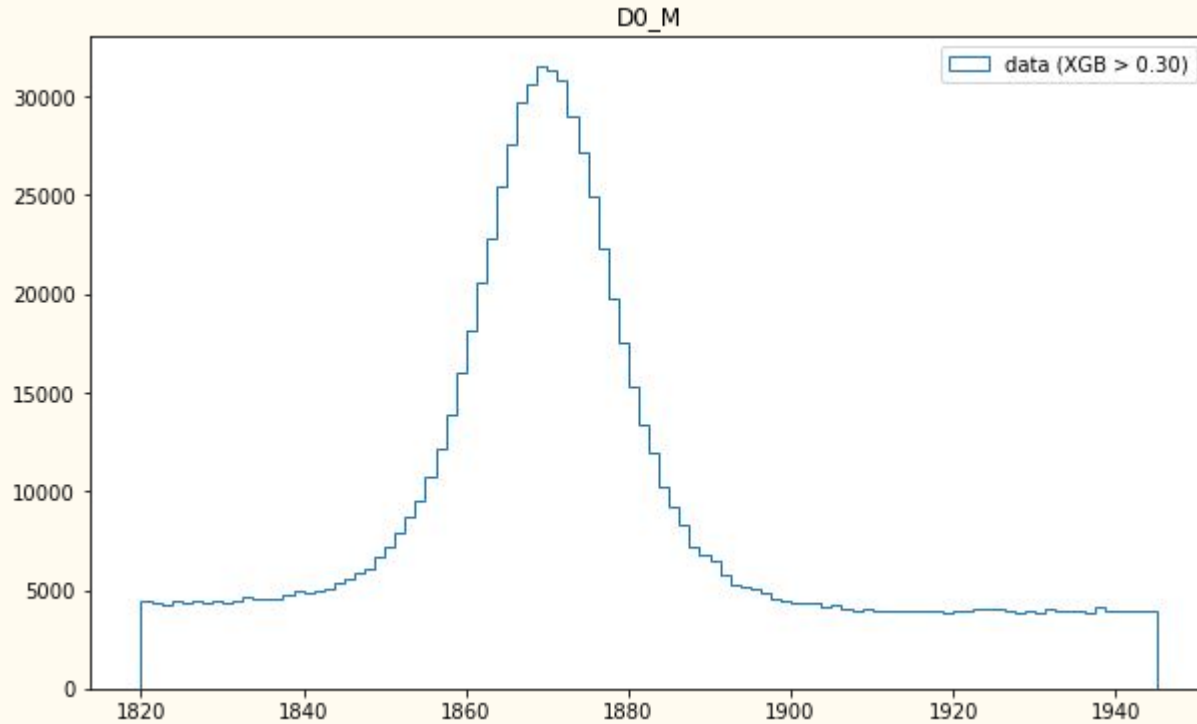
Plots for reference channel

Plots for reference channel

Plots for reference channel
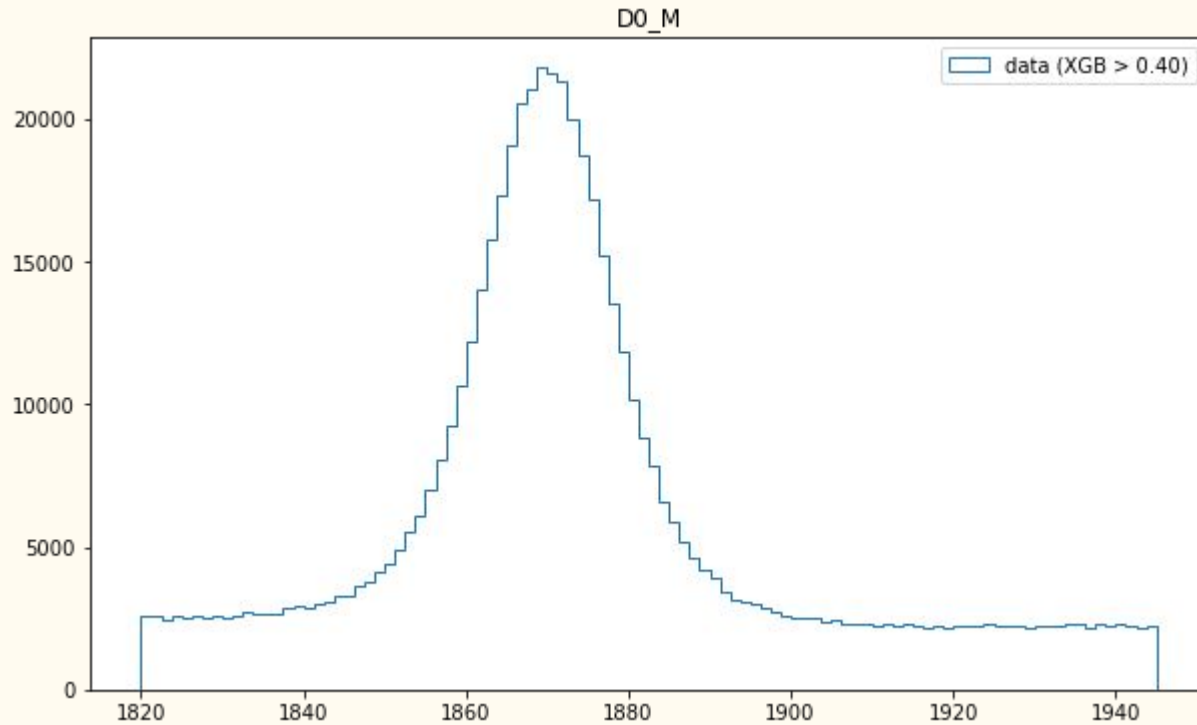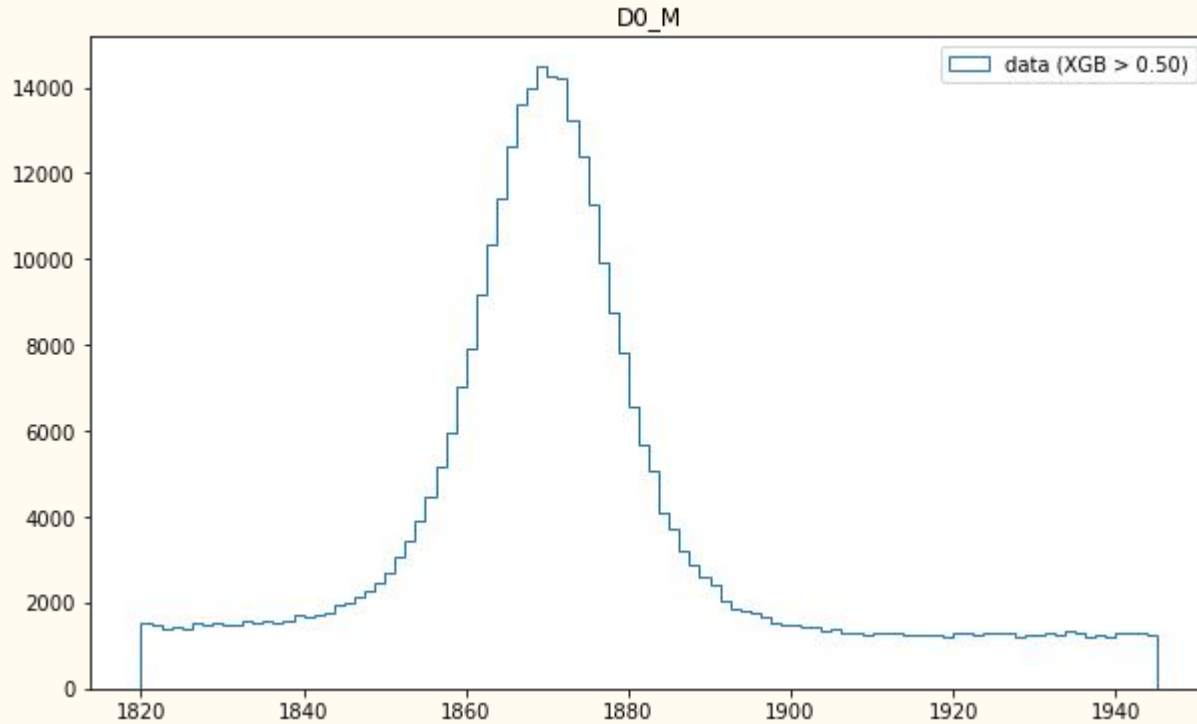

D0_M

data (XGB > 0.20)

Plots for reference channel
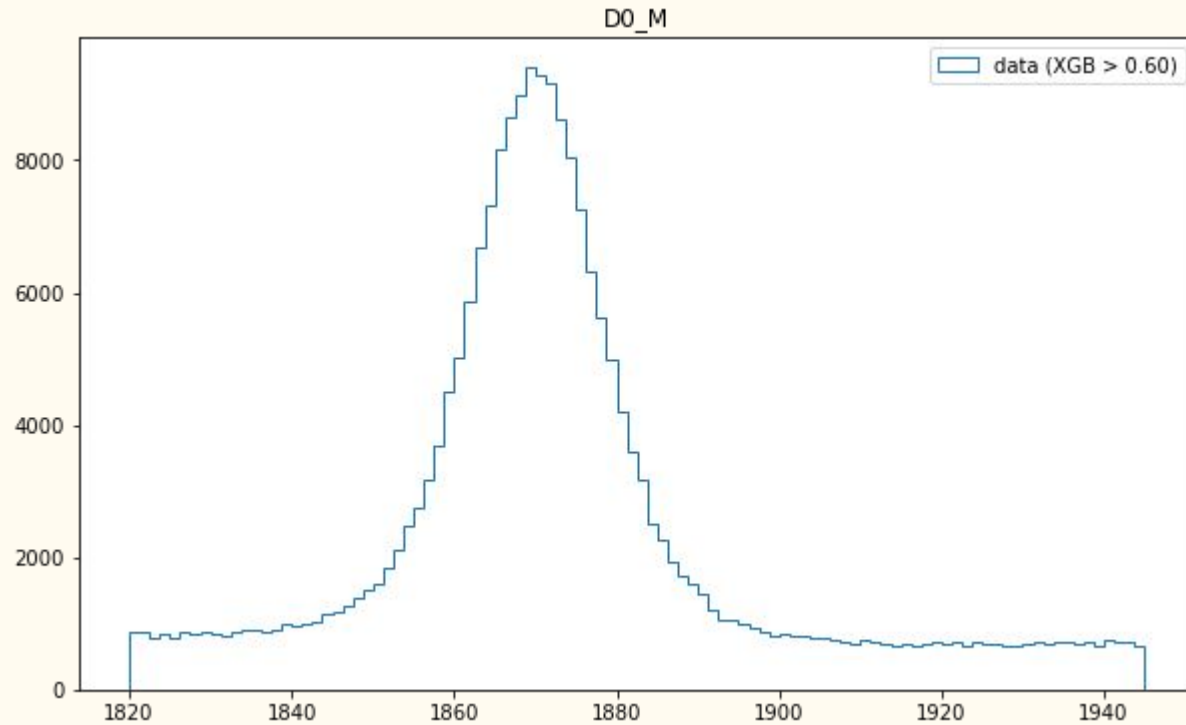
Plots for reference channel

Plots for reference channel
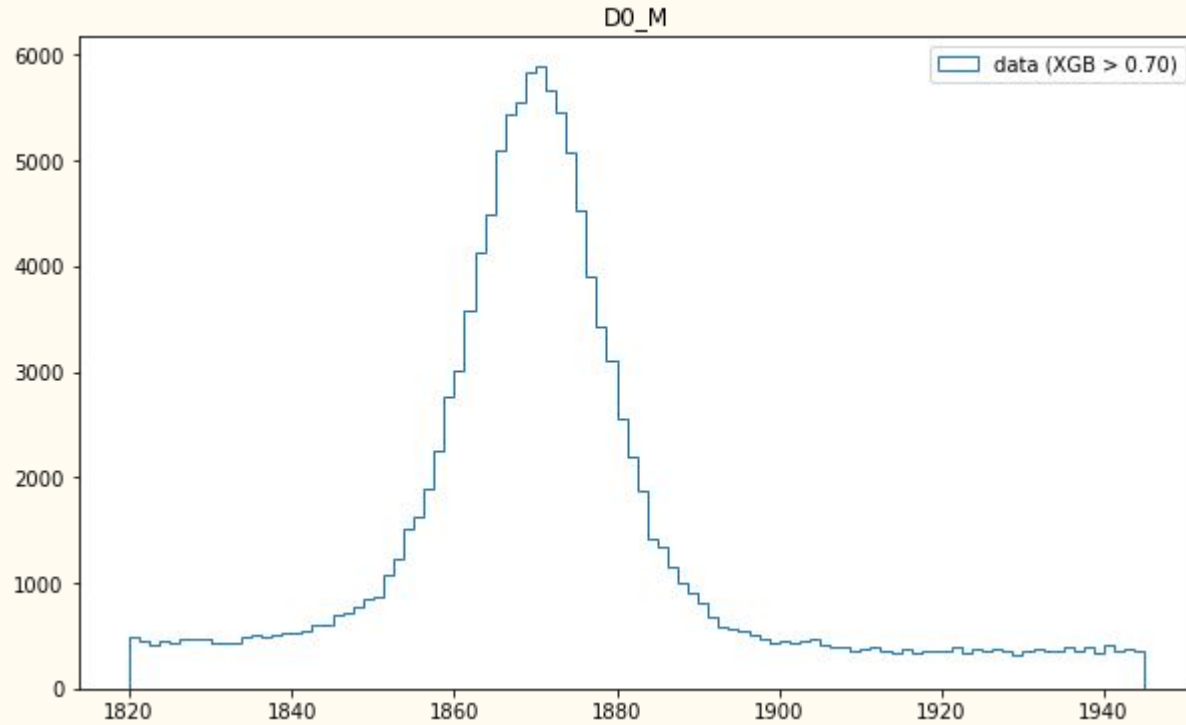
Plots for reference channel

D0_M

Plots for reference channel

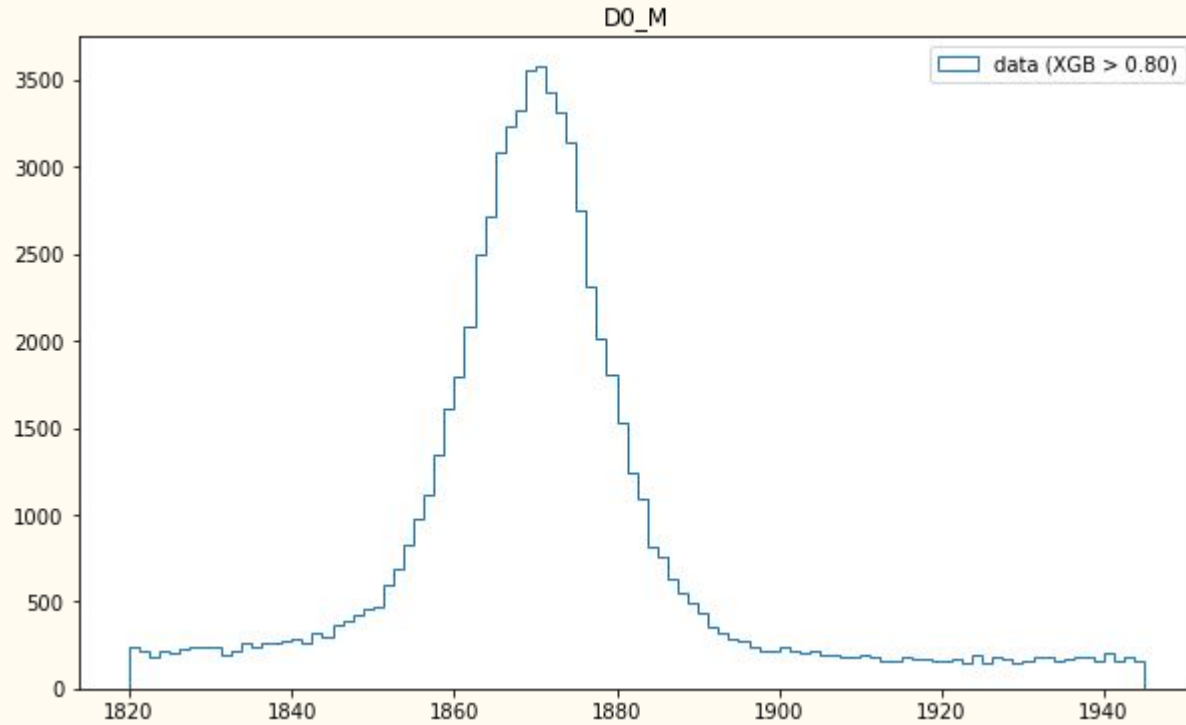# Plots for reference channel
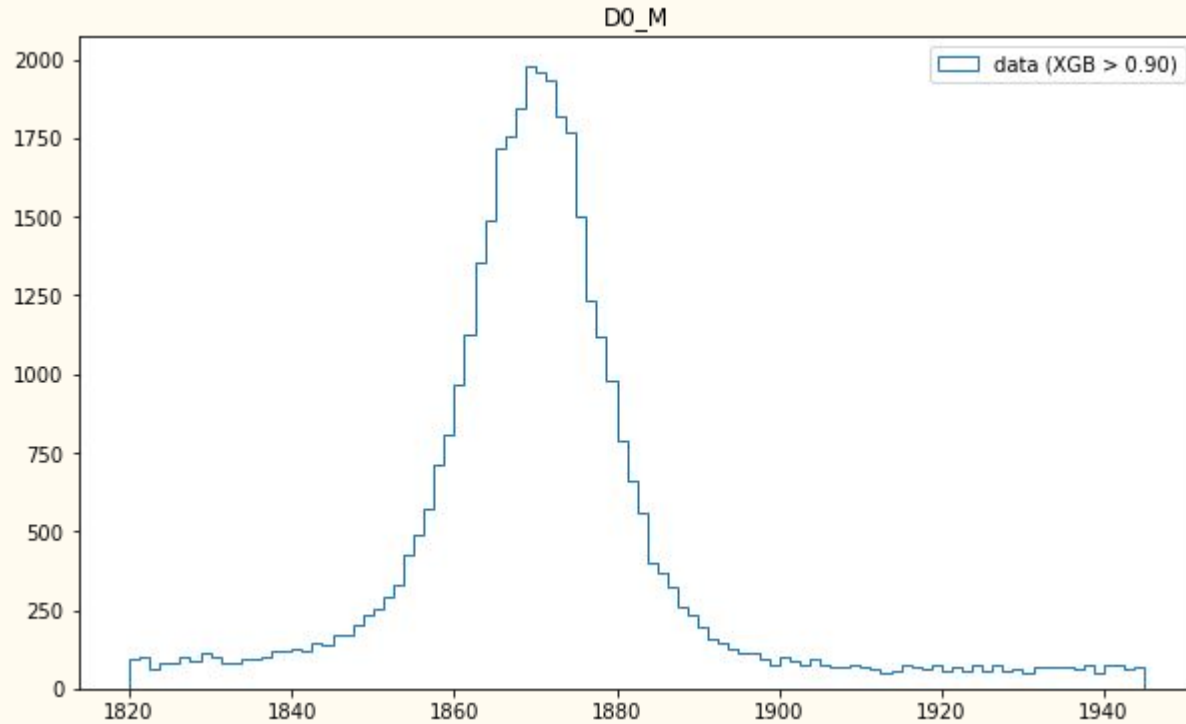
Plots for reference channel

Plots for reference channel
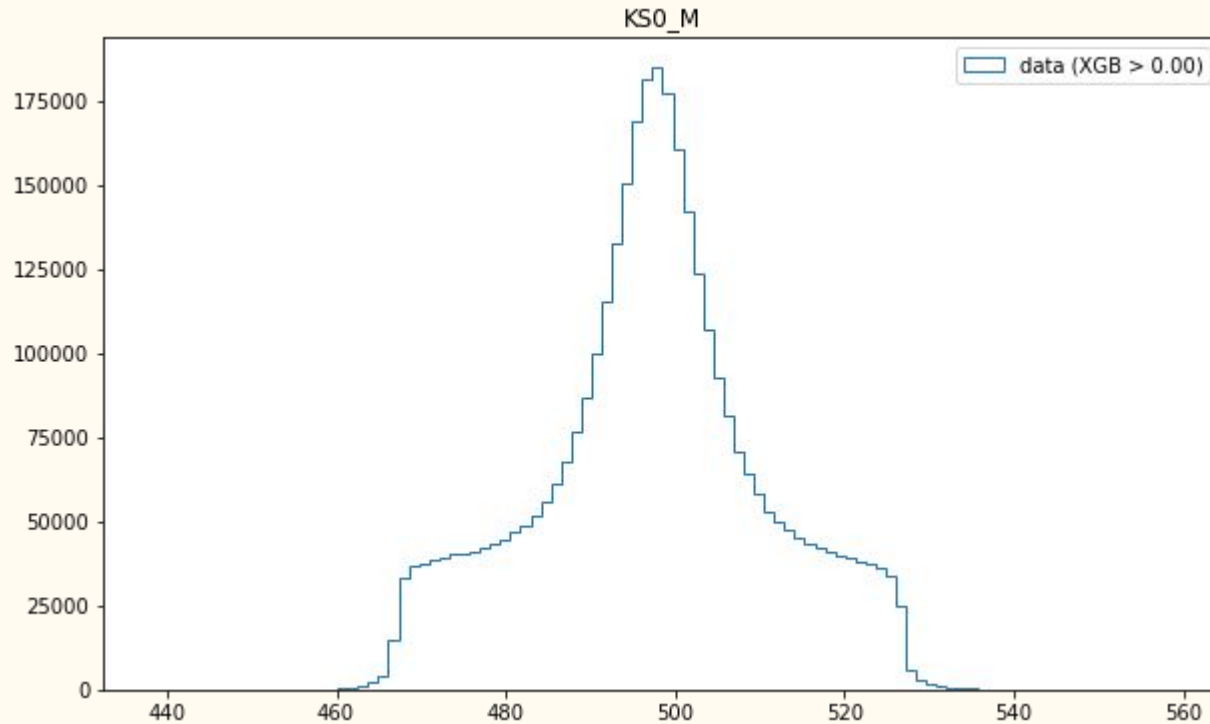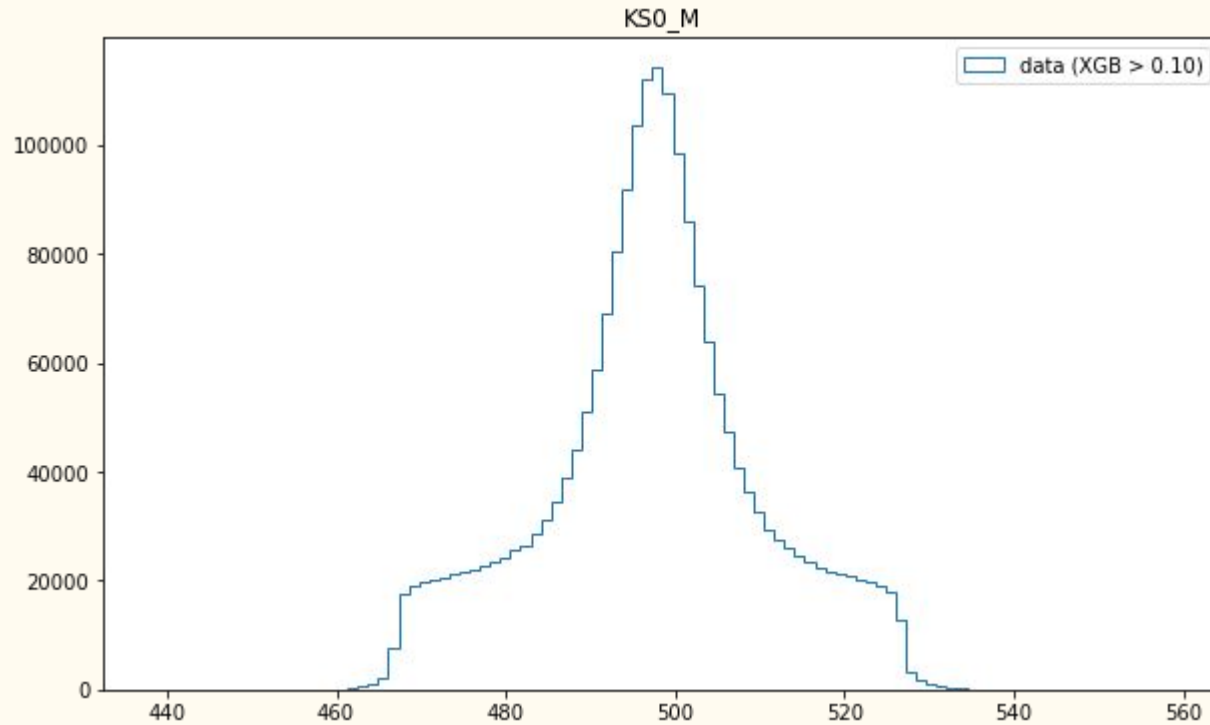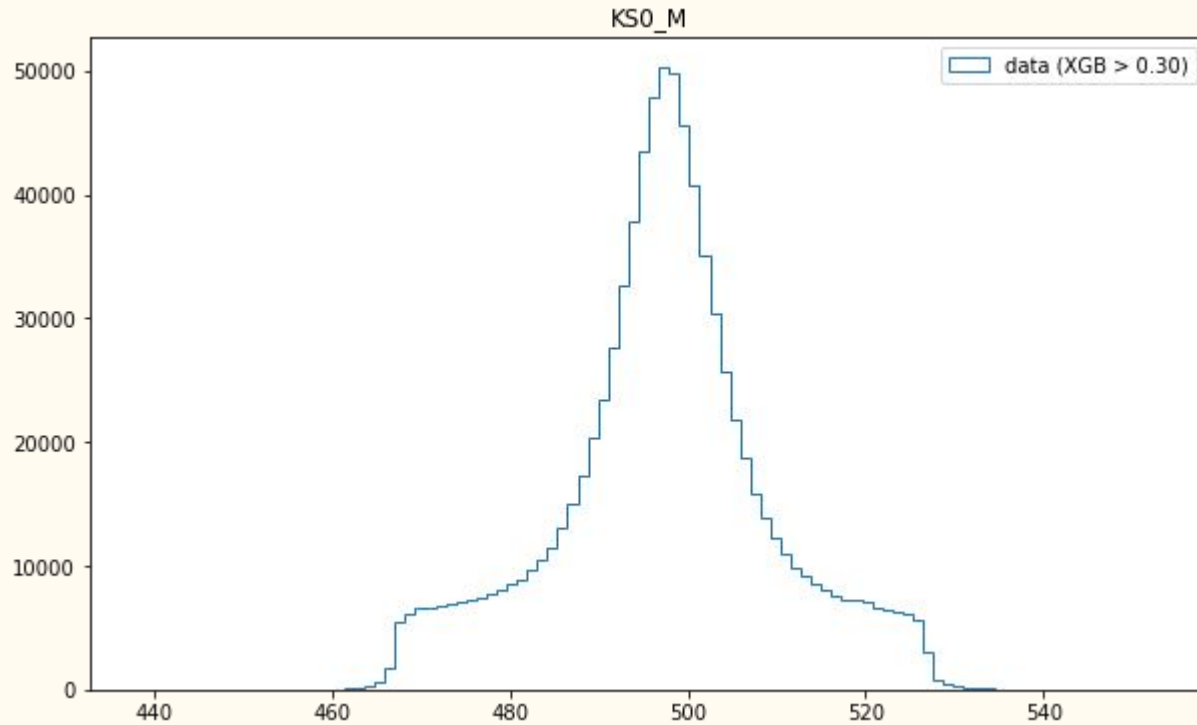
Plots for reference channel

KS0_M

Plots for reference channel

Plots for reference channel

Plots for reference channel

Plots for reference channel

KS0_M
data (XGB > 0.60)
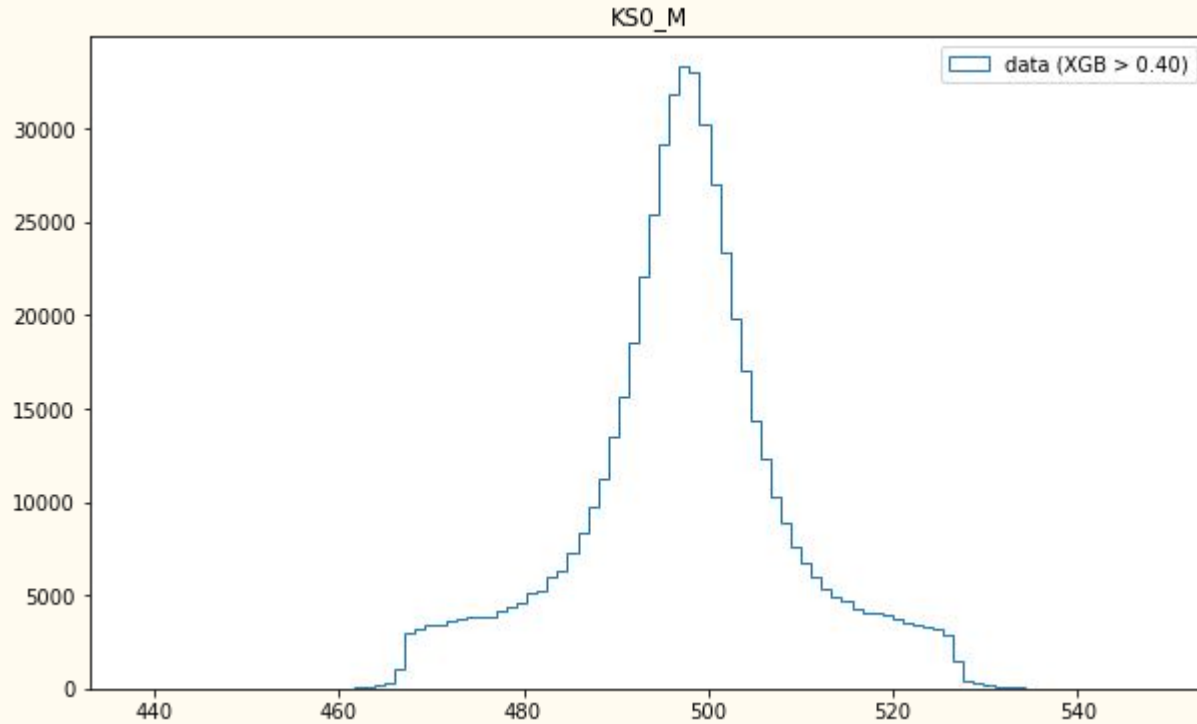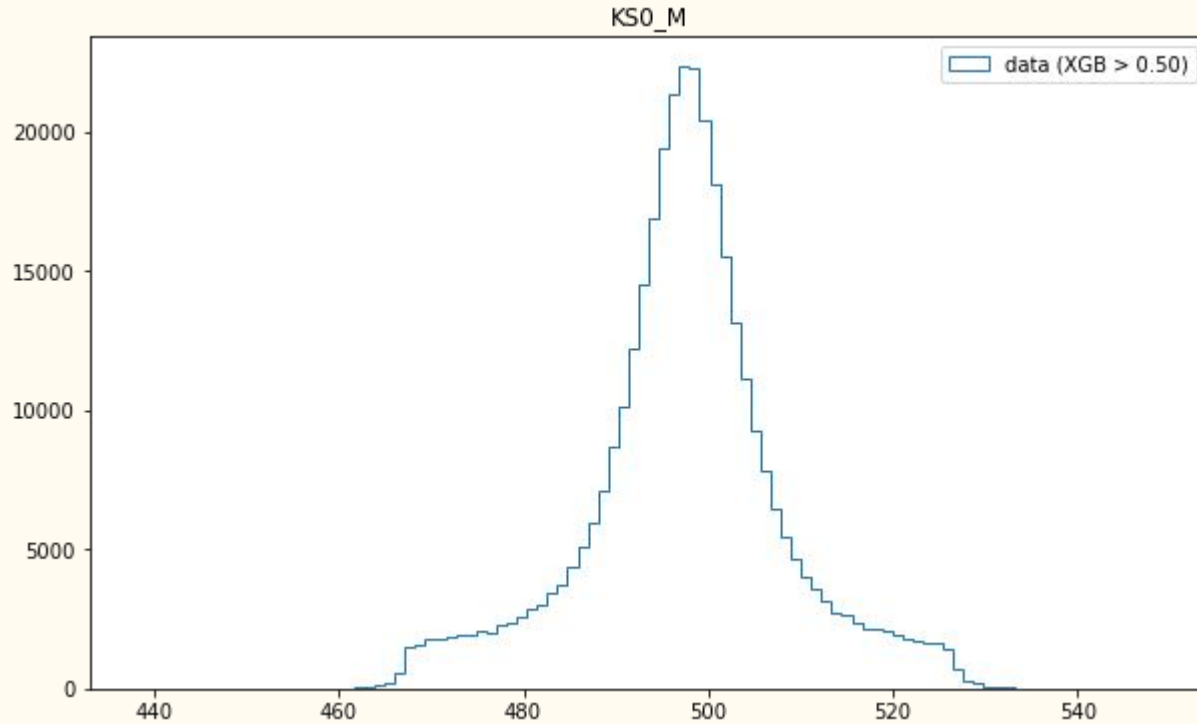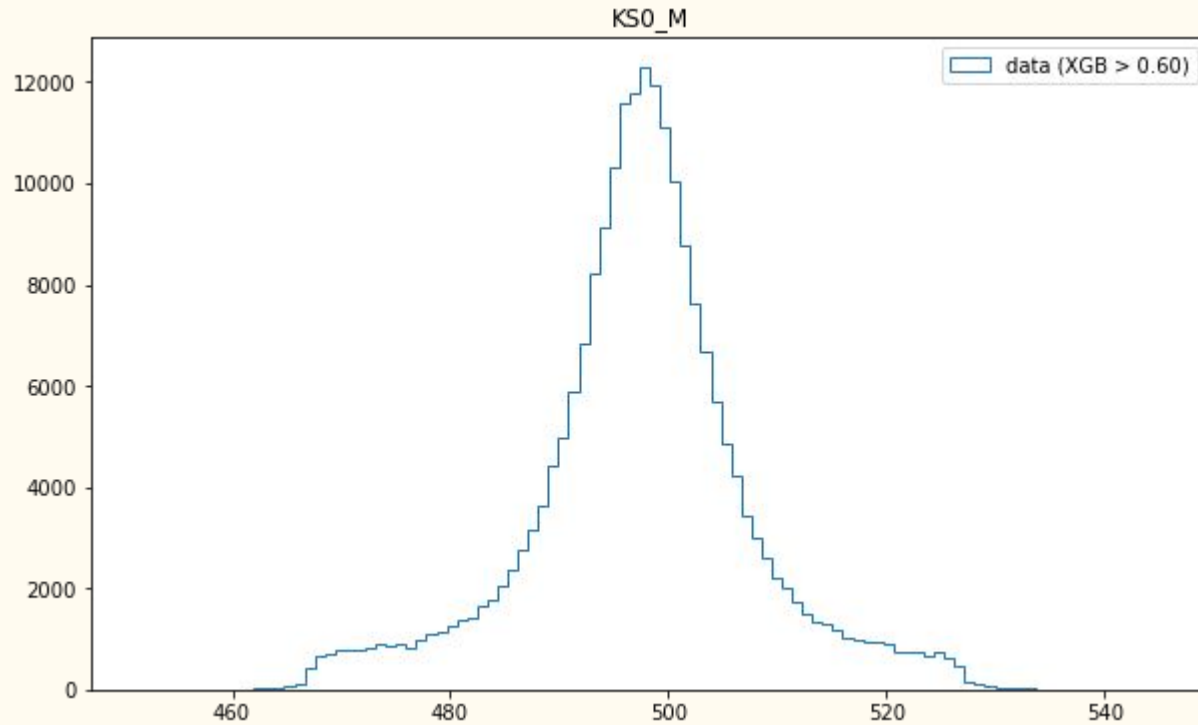
Plots for reference channel
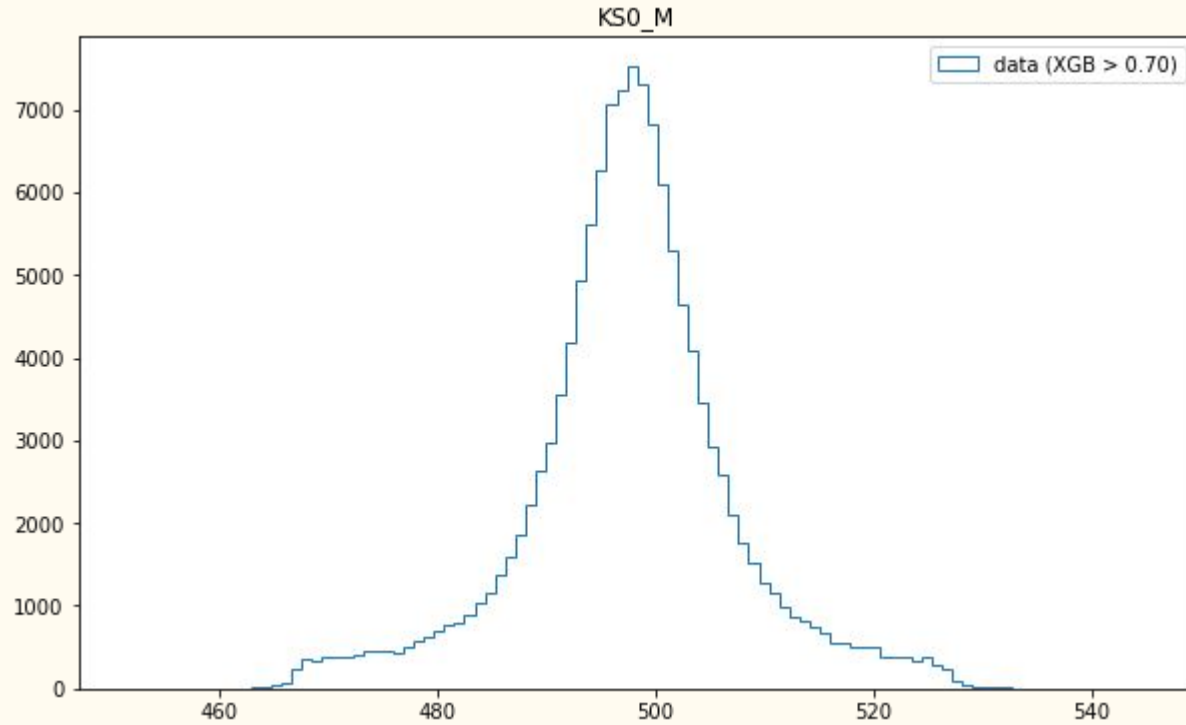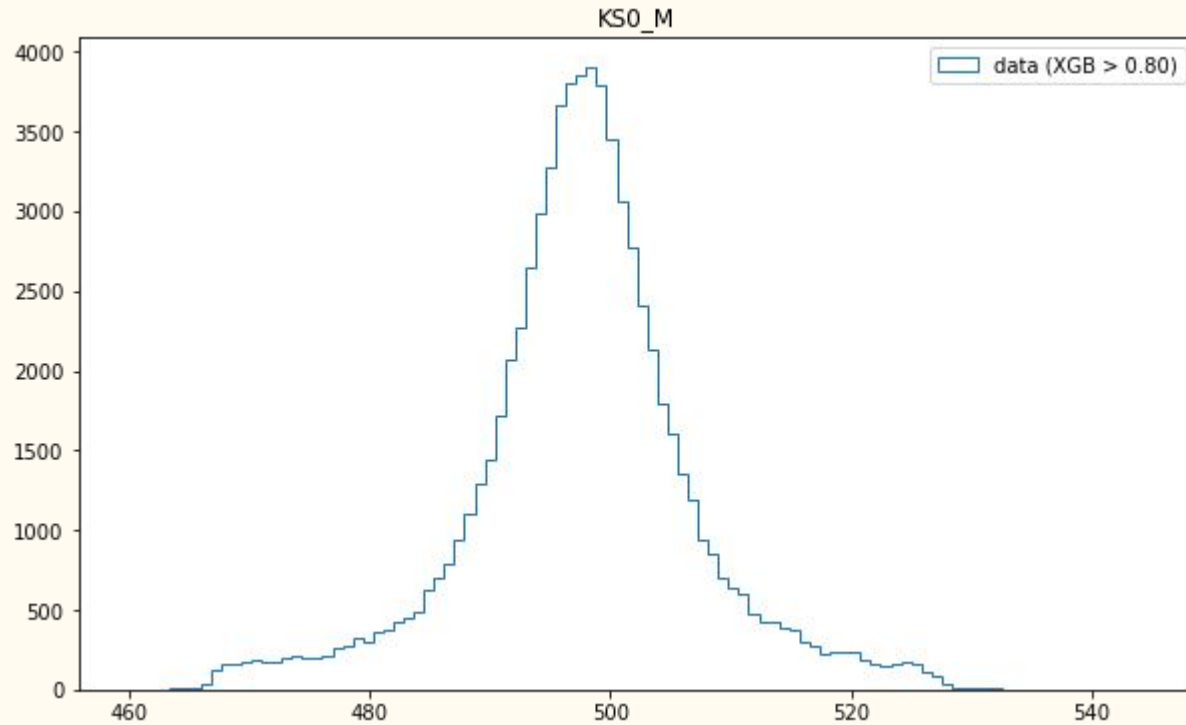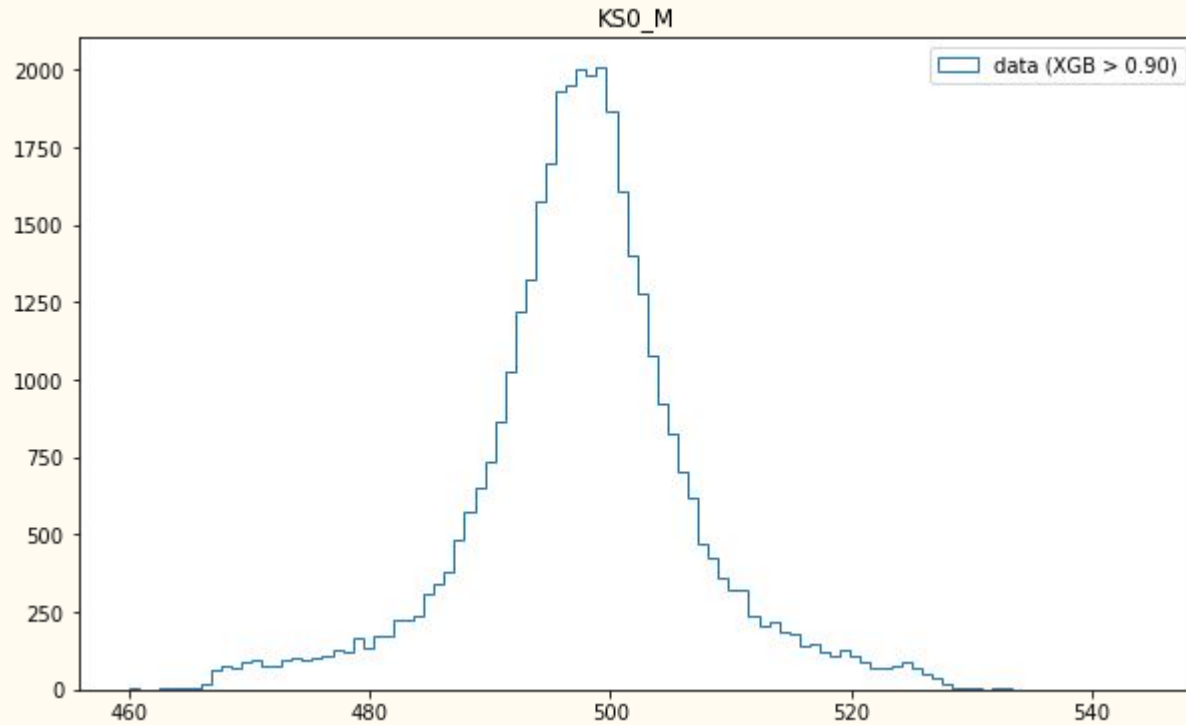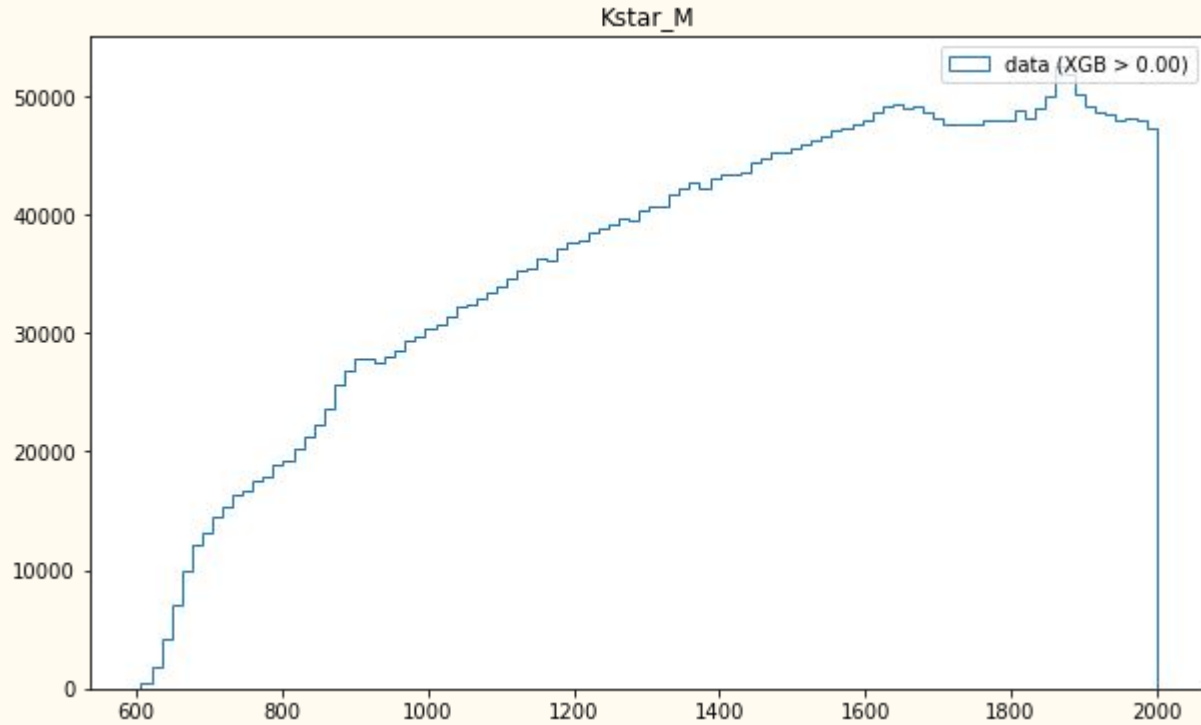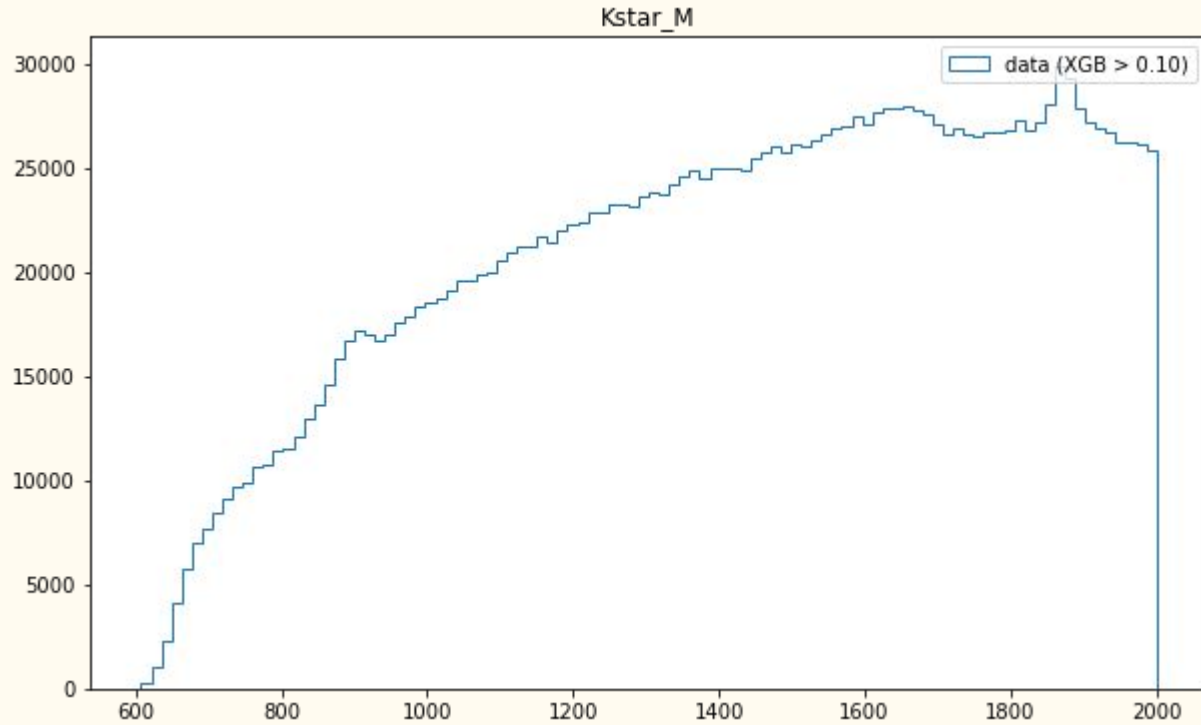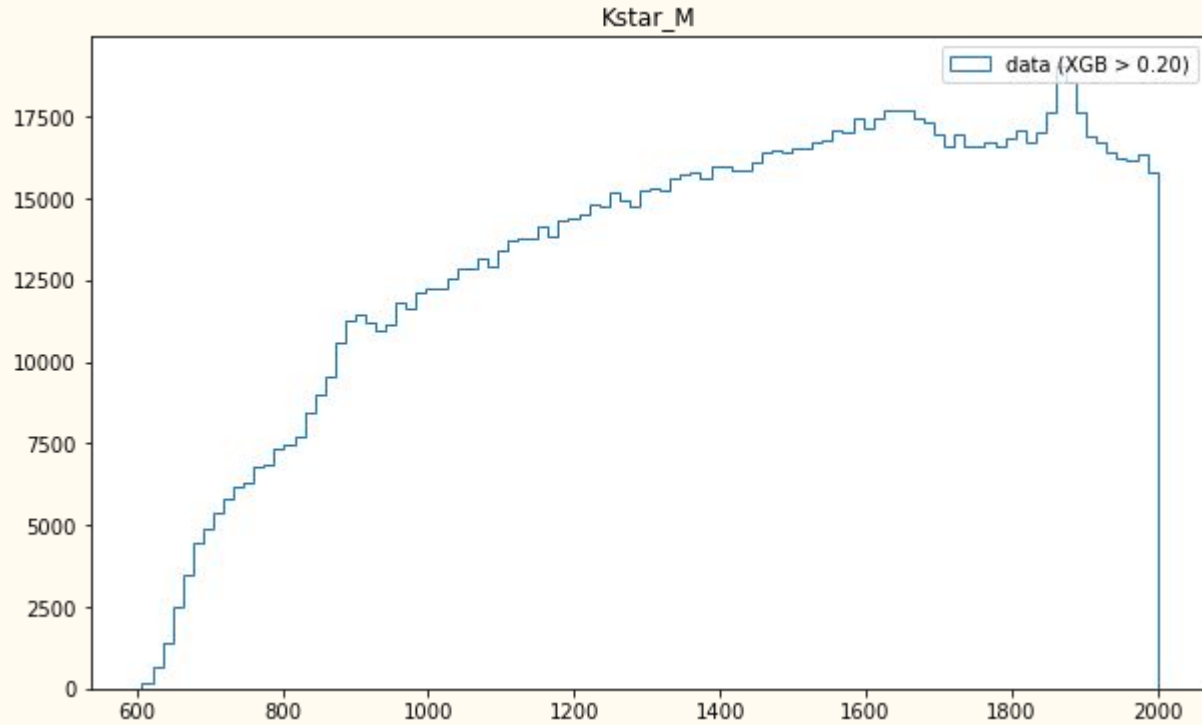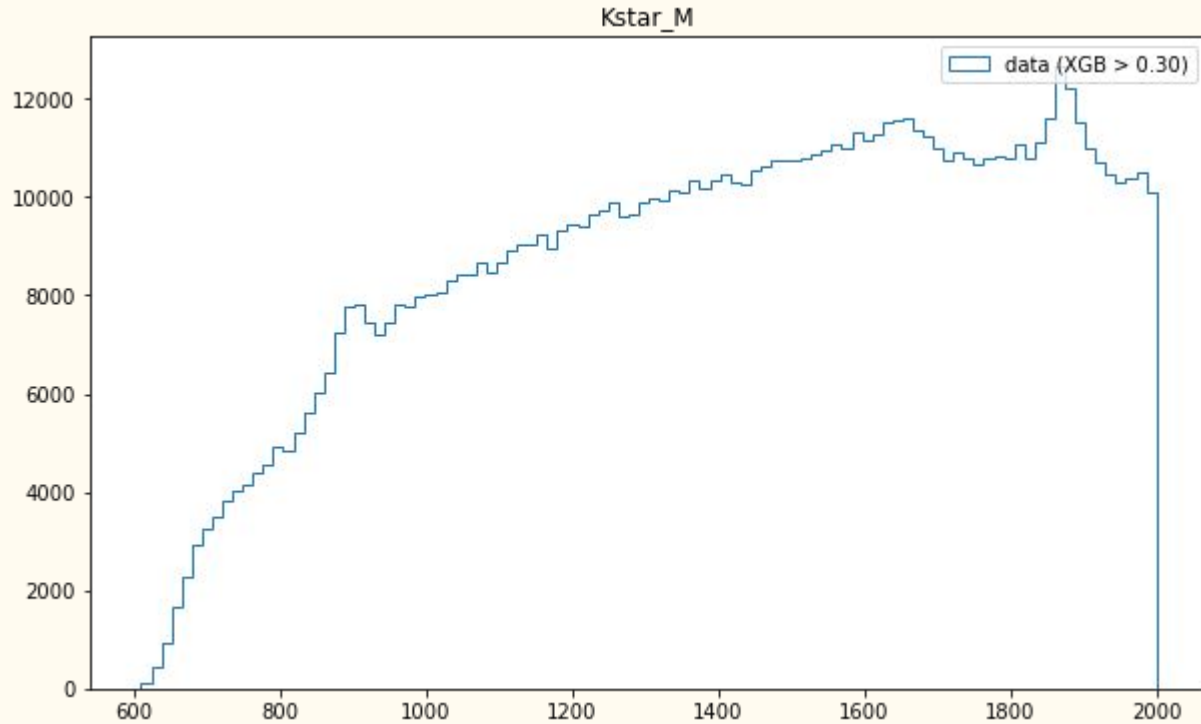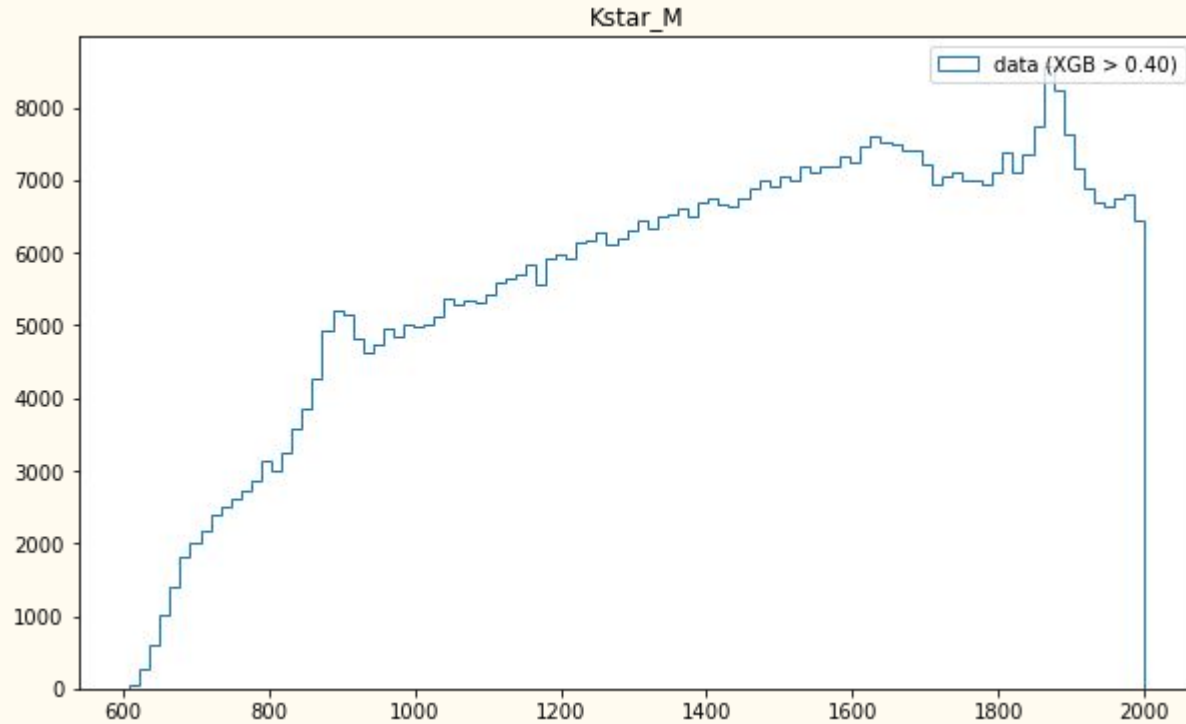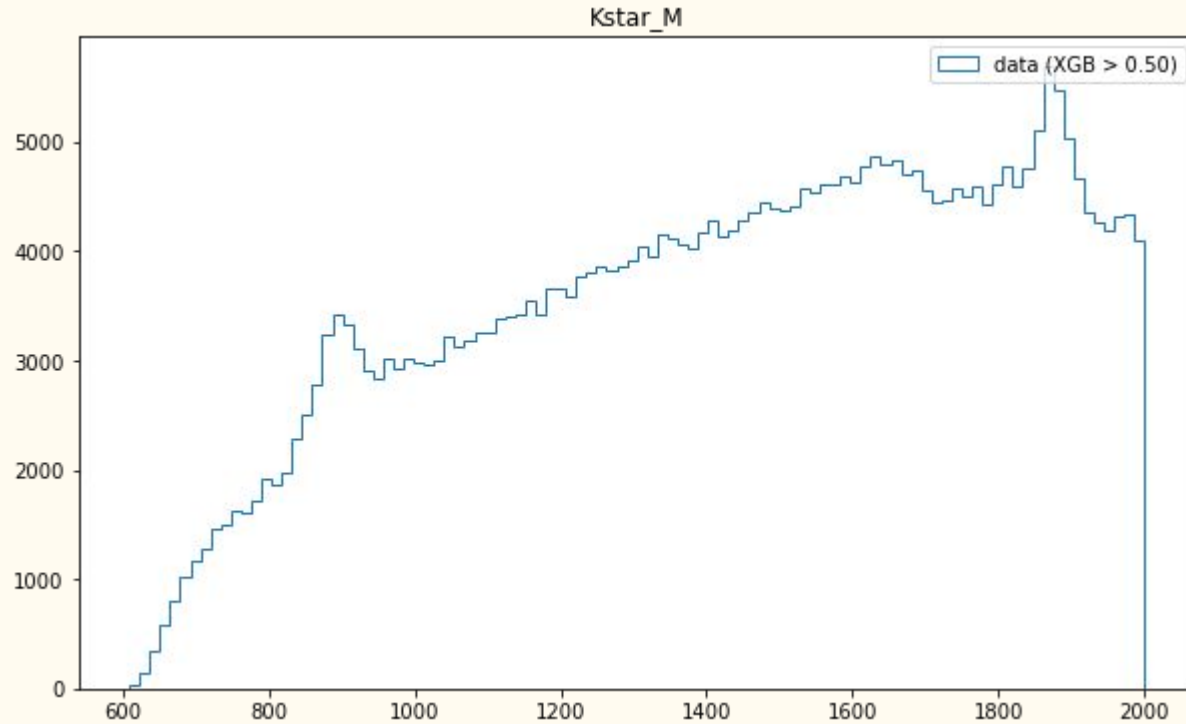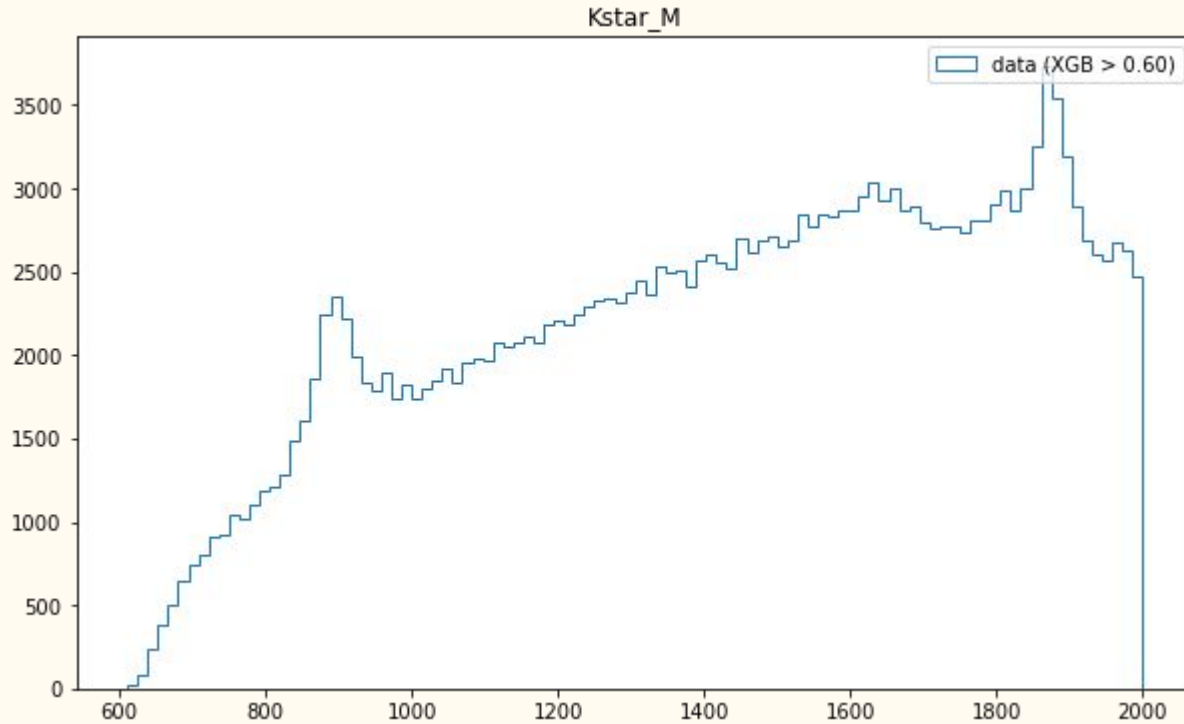

KS0_M

Plots for reference channel

Plots for reference channel

Plots for reference channel

Kstar_M

Plots for reference channel

Kstar_M

Plots for reference channel

Kstar_M

Plots for reference channel

Plots for reference channel

Kstar_M

Plots for reference channel

Kstar_M

Plots for reference channel
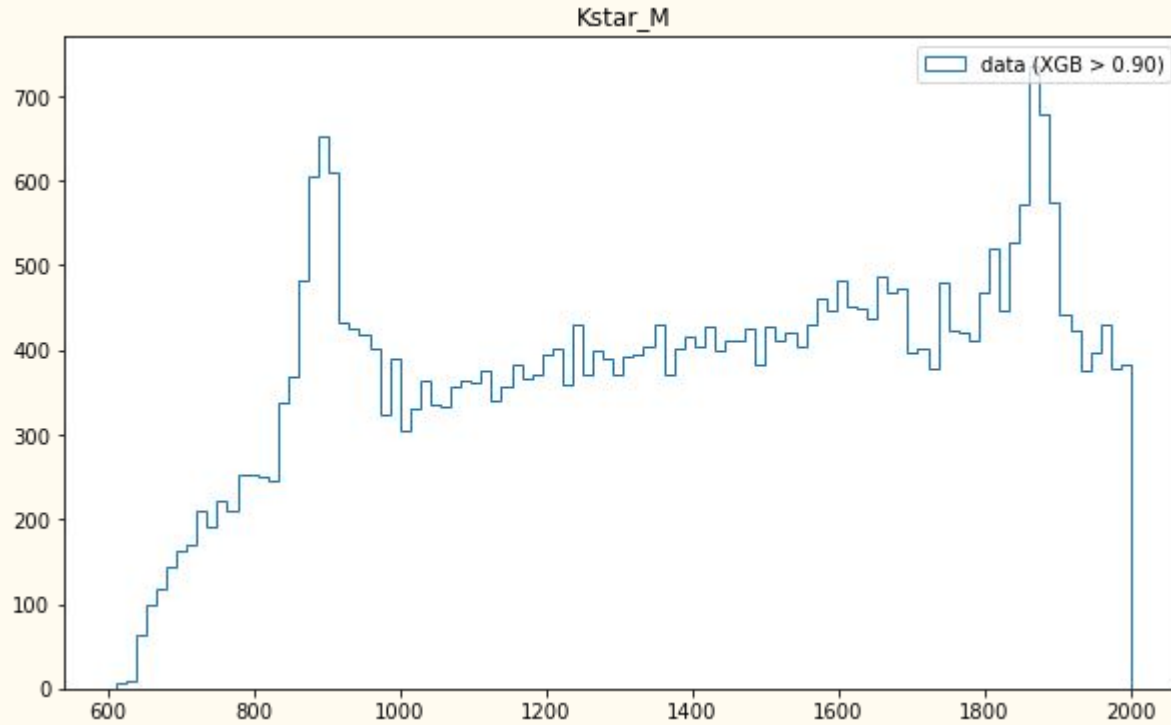
Plots for reference channel

Kstar_M

Plots for reference channel

Kstar_M

Plots for reference channel

Kstar_M

Summary

The pipeline was successfully applied to control channel and it shows very good performance. Next thing to do is application to signal channel.

Questions?

Thank you for your attention