Angular Analysis of the $B^0 \rightarrow K^{*0}(892)\,\mu^+\mu^-$ Decay Channel

Using Data from CMS Experiment at $\sqrt{s}$ = 13.6 TeV

Rare Decays Meeting

Student: Thiago de Andrade Rangel Monteiro
Advisors: Eliza Melo & Gilvan Alves

april 13 of 2026

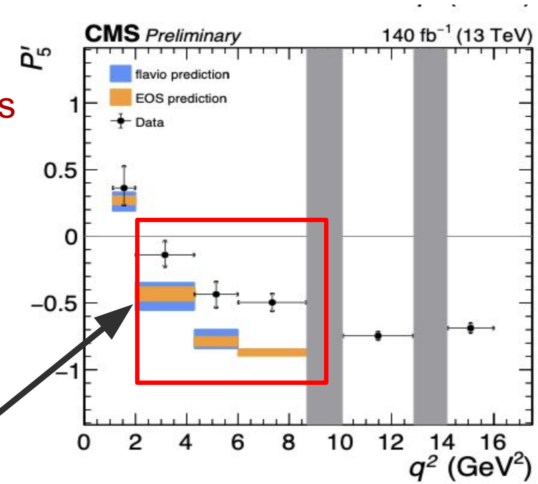
Flavour Anomalies $b \to s\ell^+\ell^-$

FCNC process in SM -- BF$\lesssim 10^{-5}$

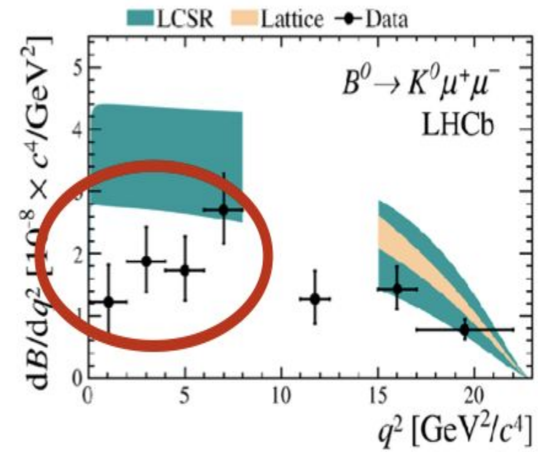
- ❑ Highly sensitive to New Physics
- ❑ (indirect tests of NP @TeV)
  - SUSY
  - Lepto Quarks
  - Heavy gauge bosons

From Feynman Diagrams to the exp. observables:

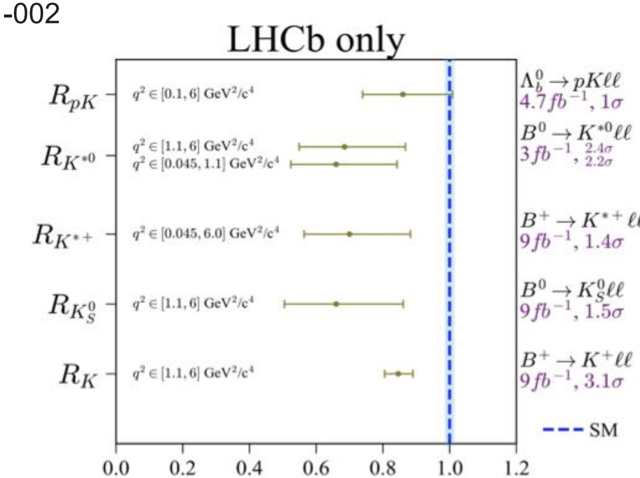
- ❑ Branching Fraction
- ❑ Diferencial Decay Rate
- ❑ **Angular Observables**

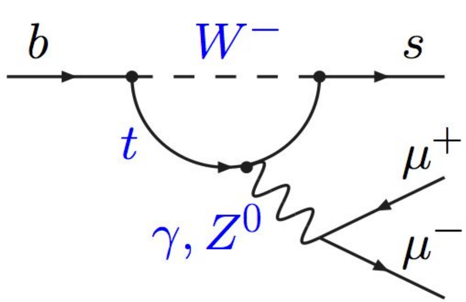


CMS PAS BPH-21-002

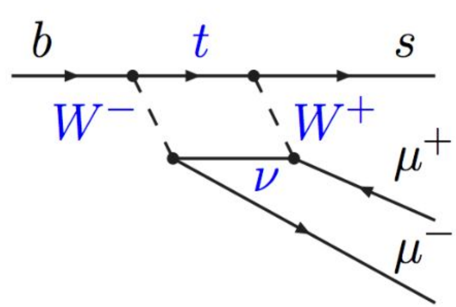


Flavor 2019 | Anomalies in b->sll transitions



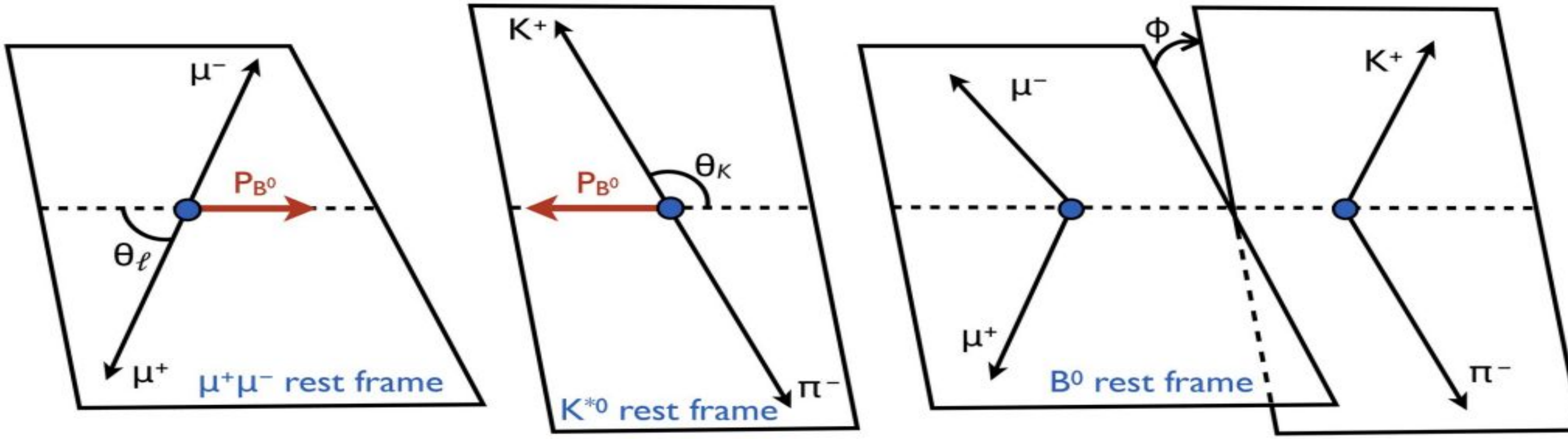


Feynman Diagram of penguin type



Feynman Diagram of box type

Taxa de Decaimento Angular



This decay is fully described by four kinematic parameters: the squared invariant mass of the dimuon system (q^2) with the latter three being angular parameters related to the decay topology, given by:

- θ_K: The angle between the K^+ direction and the direction opposite to the B^0, in the K^{*0} rest frame.
- θ_L: The angle between the μ^+ direction and the direction opposite to the B^0 in the dimuon rest frame ($\mu^+\mu^-$).
- ϕ: The azimuthal angle between the decay planes of the hadronic ($K\pi$) and leptonic ($\mu\mu$), systems in the B^0 rest frame.

DataSets and Pre-selections

For this analysis we wiil be using the run III datasets, but for this presentation we will show the results for 2022 Datasets.

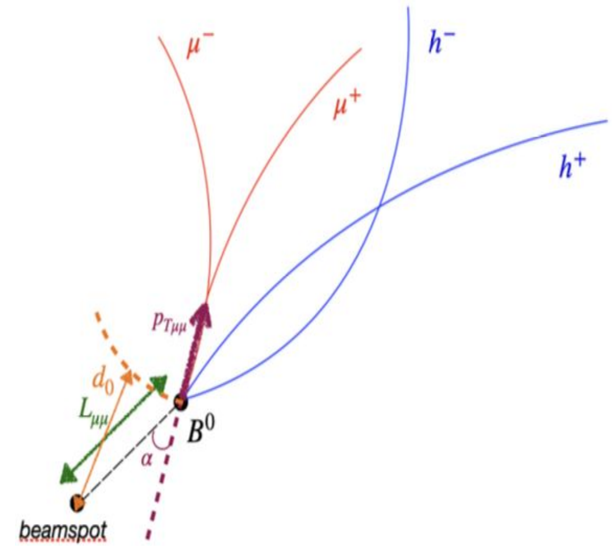
Era	Official Dataset Path	Luminosity [fb^{-1}]
C	/ParkingDoubleMuonLowMass[0-7]/Run2022C-PromptReco-v1/MINIAOD	5.01
D	/ParkingDoubleMuonLowMass[0-7]/Run2022D-PromptReco-v1/MINIAOD	2.67
E	/ParkingDoubleMuonLowMass[0-7]/Run2022E-PromptReco-v1/MINIAOD	5.23
F	/ParkingDoubleMuonLowMass[0-7]/Run2022F-PromptReco-v1/MINIAOD	16.00
G	/ParkingDoubleMuonLowMass[0-7]/Run2022G-PromptReco-v1/MINIAOD	2.78
Total		**31.69**

MC sampes:

Sample	Dataset Path	Events	σ [pb]	BF
Signal (2022)	/BdtoKstar2Mu_.../Run3Summer22MiniAODv4-FilterFix.../MINIAODSIM	5,839,635	7.20×10^6	6.07×10^{-7}
Signal (2022EE)	/BdtoKstar2Mu_.../Run3Summer22EEMiniAODv4-FilterFix.../MINIAODSIM	19,906,620	7.21×10^6	6.07×10^{-7}
Control J/ψ (2022)	/BdtoJpsiKstar_.../Run3Summer22MiniAODv4.../MINIAODSIM	3,257,972	13.80×10^6	5.07×10^{-5}
Control J/ψ (2022EE)	/BdtoJpsiKstar_.../Run3Summer22EEMiniAODv4.../MINIAODSIM	11,471,877	13.80×10^6	5.07×10^{-5}
Control $\psi(2S)$ (2022)	/BdtoKstarPsi2s_.../Run3Summer22MiniAODv4.../MINIAODSIM	3,346,531	16.88×10^6	3.16×10^{-6}
Control $\psi(2S)$ (2022EE)	/BdtoKstarPsi2s_.../Run3Summer22EEMiniAODv4.../MINIAODSIM	12,081,650	16.88×10^6	3.16×10^{-6}

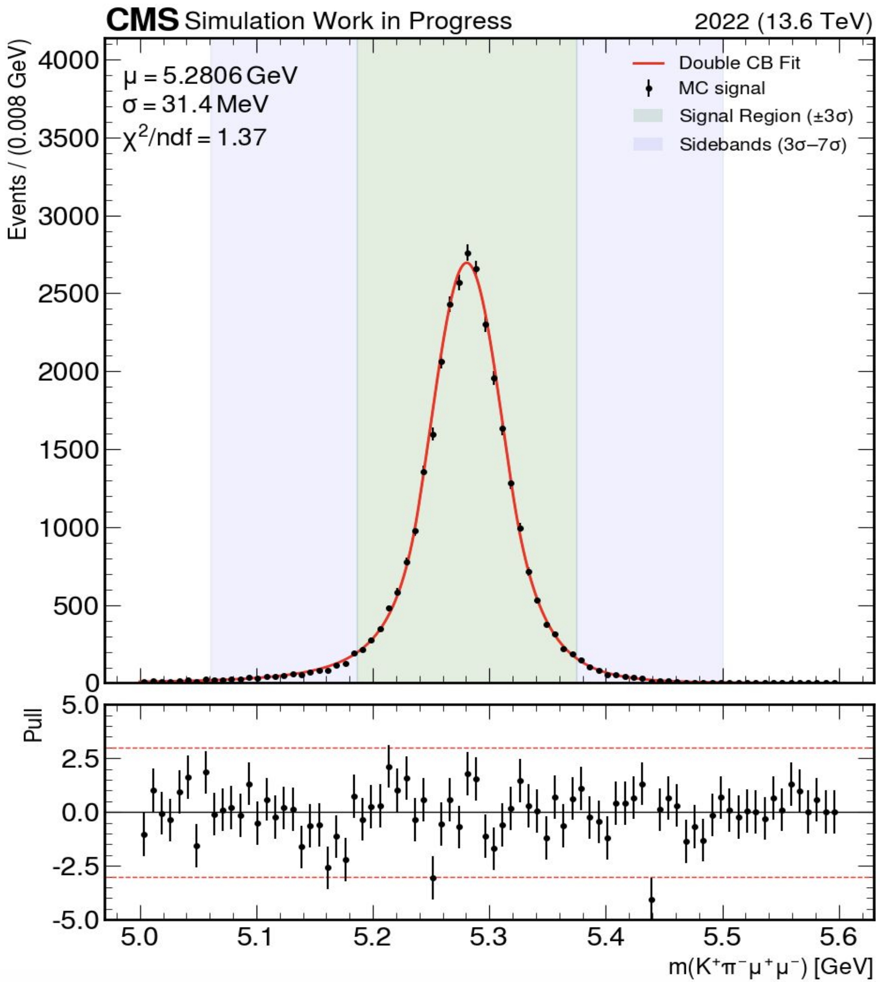
DataSets and Pre-selections

Variable	Cut	Description
$p_T(\mu\mu)$	> 6.9 GeV	Minimum transverse momentum of the muon pair.
$P(\chi^2_{vtx})$	> 0.1	Dimuon vertex fit probability.
$L_{xy}/\sigma_{L_{xy}}$	> 3	Transverse flight distance significance of dimuon.
$\cos(\alpha_{2D})$	> 0.9	Alignment between the B momentum and the vertex.
$p_T(\mu_{1,2})$	> 4 GeV	Transverse momentum of each individual muon.
$\lvert\eta(\mu_{1,2})\rvert$	< 2.4	Acceptance of the muon spectrometer system.
Soft MVA Run 3	> 0.74	Muon identification for low-energy signals.
$p_T(trk_1)$	> 2.0 GeV	Minimum transverse momentum of the K^{*0} tracks.
$\lvert\eta(trk_1)\rvert$	< 2.4	Coverage of the central tracking system.
Mass Arbitration	Eq. K^*	Choice of the $K\pi$ hypothesis closest to the true mass.
$M(K^+K^-)$	> 1.035 GeV	Contamination veto from the $B^0_s \to \phi\mu\mu$ decay.



- Furthermore, a cut on the dimuon mass is applied in the regions of (1.0 - 2.7 and 4.0 - 4.8) GeV, excluding the ressonant channels. The selections above are quite the same as those used in the Run II analysis.
- The selection includes requirements on the dimuon vertex fit quality, displacement significance, pointing angle, and the transverse momenta of the final-state particles. Additional trigger and muon identification criteria are also imposed.

Estratégia de Análise



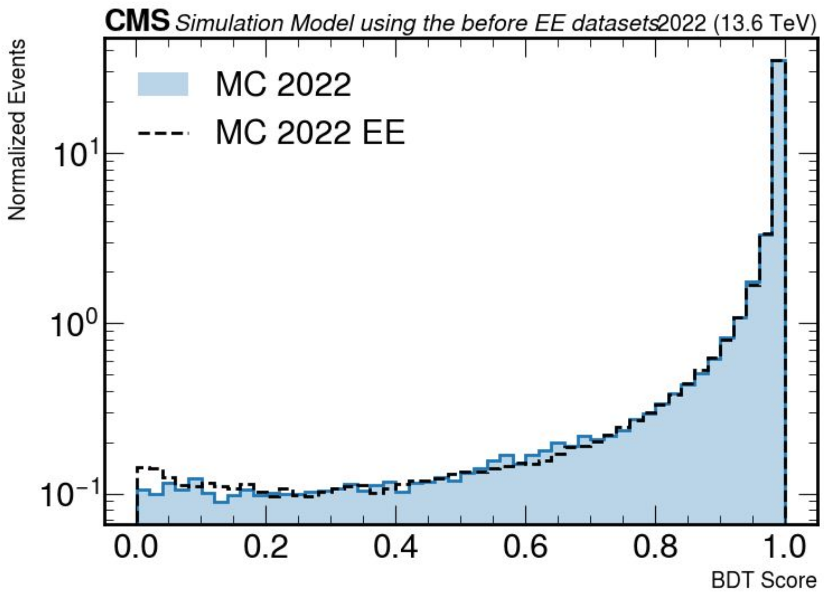
The signal region is optimized using non-resonant dimuon Monte Carlo (MC) samples.

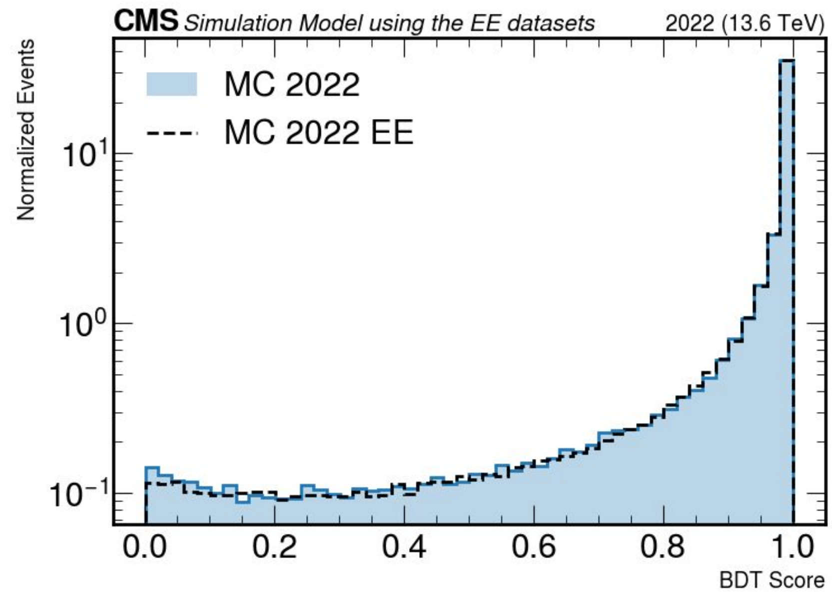
- **Signal Window:** Defined as the region within ±3σ around the B^0 mass mean $(M_{B0}$ +- $3\sigma)$.

- **Background Region (Sidebands):** Defined as the mass range between 3σ and 7σ from the mean $([M_{B0}-7\sigma, M_{B0}-3\sigma]$ e $[M_{B0}+3\sigma, M_{B0}+7\sigma])$.

Model Training: Once the signal region and sidebands are defined, we train a Boosted Decision Tree (BDT) using the XGBoost algorithm. The objective is to perform an accurate event classification, effectively separating the true physical signal from the combinatorial background

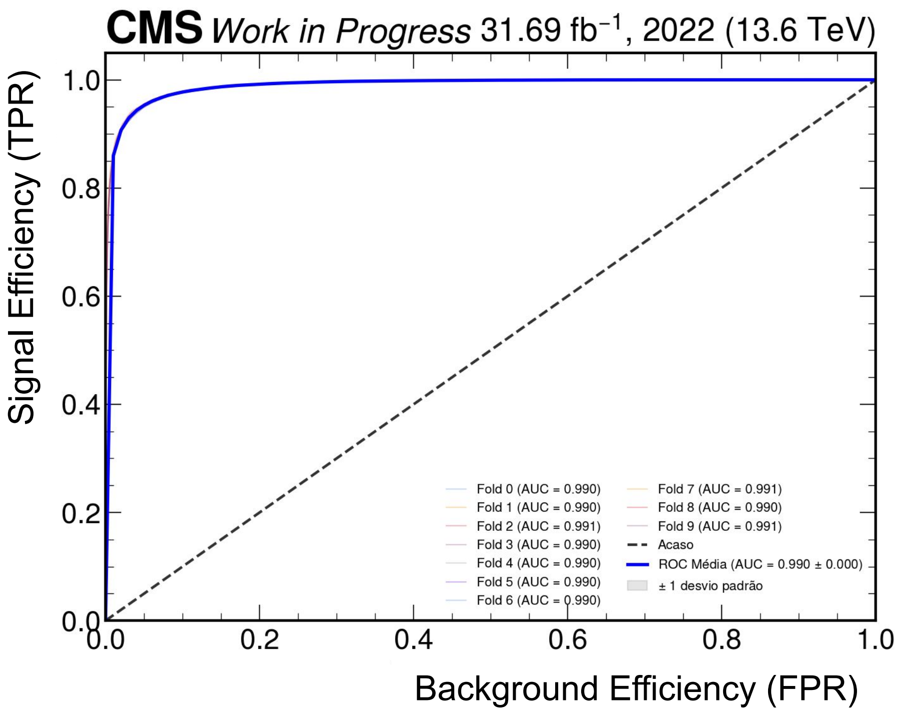
Machine Learning Analysis

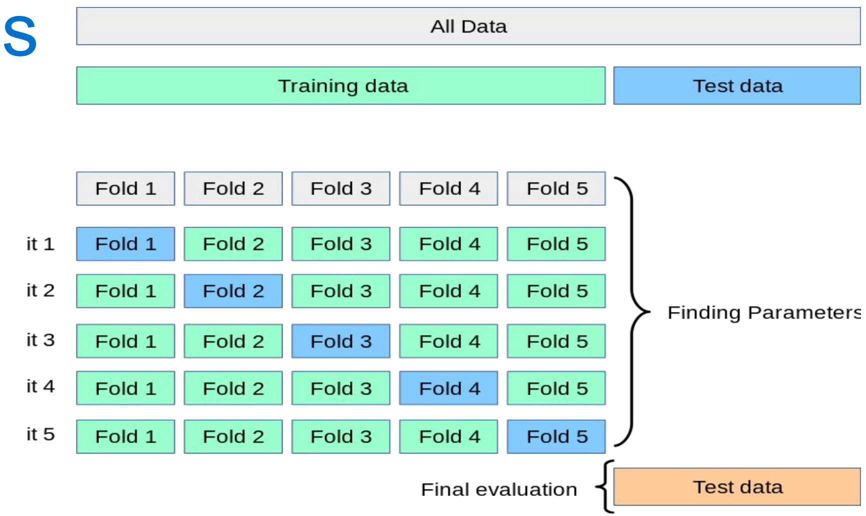




- ❑ First of all it's important to notice that 2022 dataset are divided in two eras before and after EE (ECAL Leaking), so to understand if it'' possible to analyze the full dataset as a single one we trained a BDT with XGboos implementation for each of the datasets and them apply to the MC of the other era and since the BDT output are compatible we treat the 2022 as a single dataset.

Machine Learning Analysis

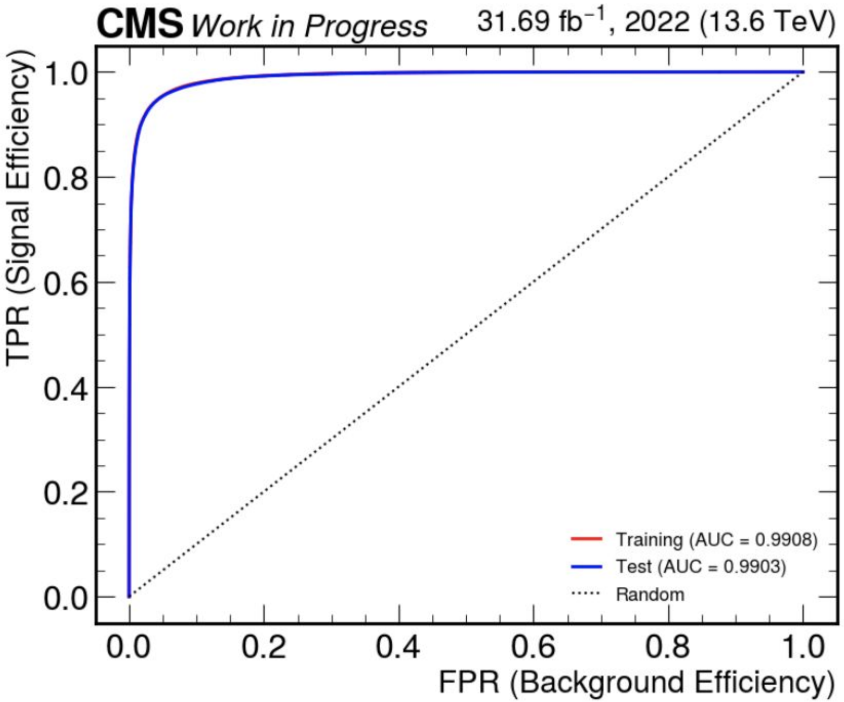


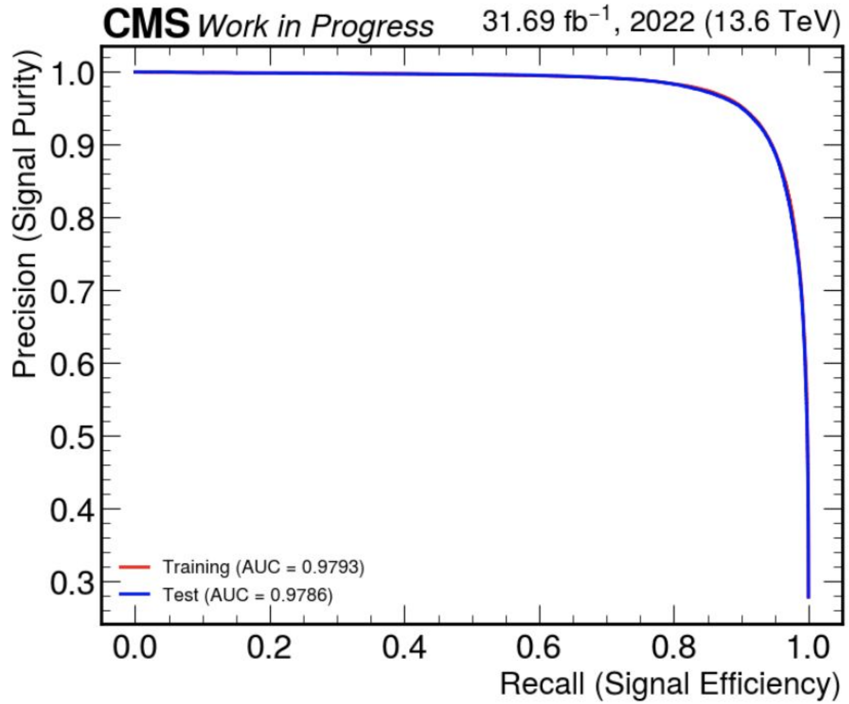


Hyperparameter	Value	Description
subsample	0.6	Fraction of samples used to train each tree.
reg_lambda	0.5	L2 regularization term on leaf weights.
reg_alpha	0.0	L1 regularization term on leaf weights.
n_estimators	500	Number of decision trees (iterations).
min_child_weight	3	Minimum sum of instance weight needed in a child node.
max_depth	3	Maximum depth of the trees.
learning_rate	0.1	Learning rate (eta) to prevent *overfitting*.
colsample_bytree	1.0	Fraction of features used per tree.

- In this graph, we present the ROC curves for each fold of our cross-validation (k-fold) process. The main point here is to observe how the curves overlap consistently, keeping the Area Under the Curve (AUC) **practically unchanged**. This stability proves that our **XGBoost** algorithm is robust and generalizes the data well; in other words, we do not suffer from *overfitting* in any specific Monte Carlo subset.

Machine Learning Analysis

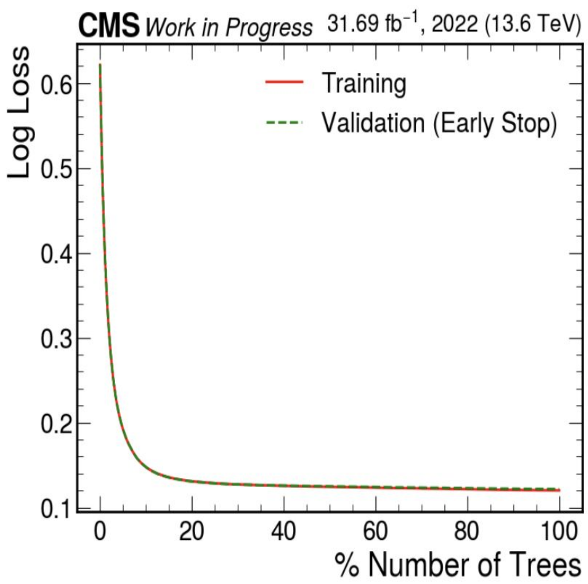


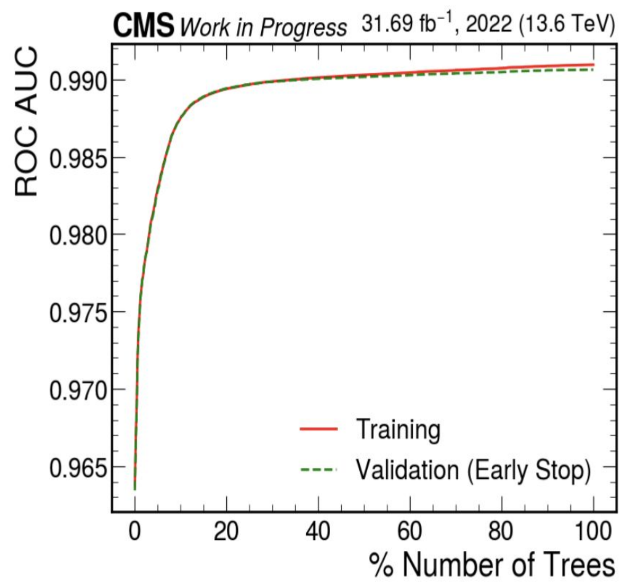


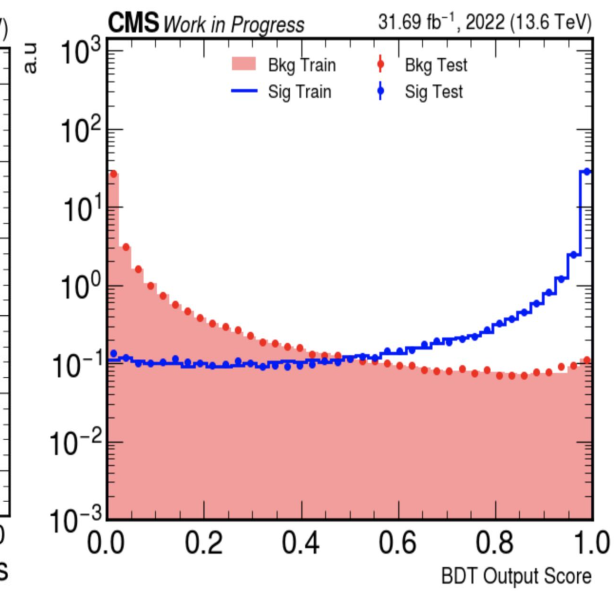
- **ROC Curve Analysis:** Demonstrates excellent discrimination power between Signal and Background. Furthermore, there is a strong agreement between the Training and Testing sets (overlapping curves), indicating the absence of overtraining.

- **Precision-Recall (PR) Curve:** The shape of the curve indicates a more stable performance in the high-precision regime, validating the quality of the current MC simulation.

Machine Learning Analysis

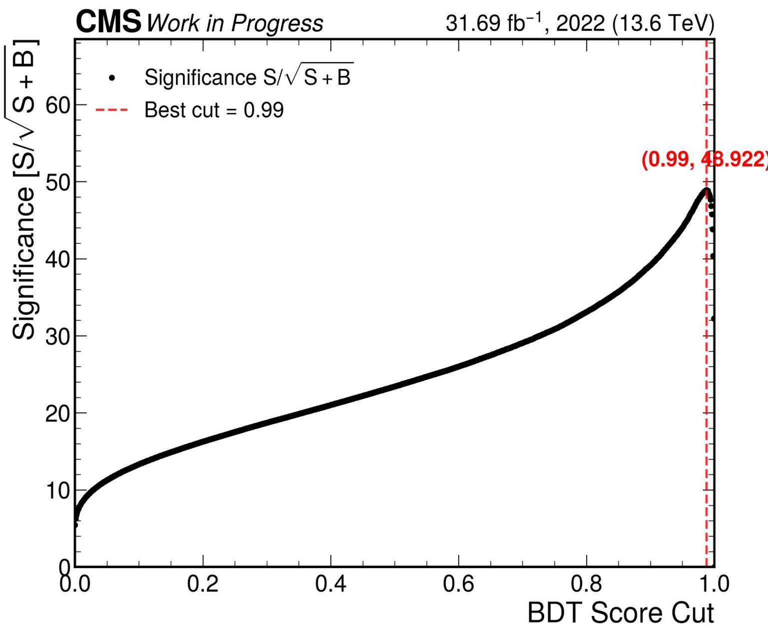


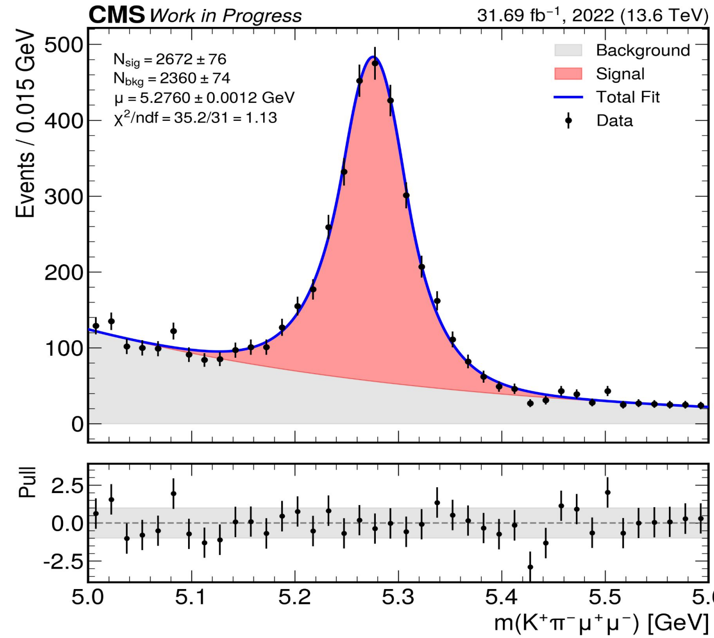




- ❏ It can be observed that the BDT score distributions for the training and testing sets strongly overlap. This excellent agreement demonstrates that the model generalizes well, effectively ruling out overtraining.
- ❏ Furthermore, the evolution of the Loss and AUC curves as a function of the number of trees shows no visible deviation between the training and testing samples.

Machine Learning Analysis



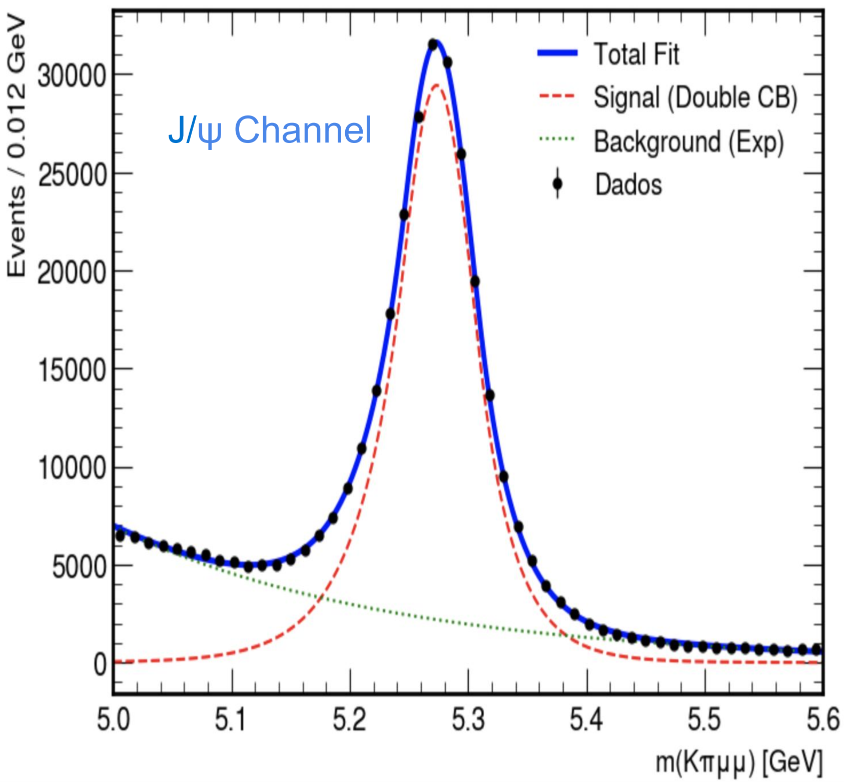


- ❑ The optimization metric is calculated using the expected Signal (S) and Background (B) yields, which are previously normalized and scaled to the target luminosity of our analysis.

- ❑ The process is performed through a continuous scan across the entire response range of the BDT algorithm. The maximum point obtained on this curve determines the final selection threshold, ensuring the best possible separation between signal events and the combinatorial background.

MC Correction: Reweighting Procedure

To compare the data with MC we use the resonant like J/ψ and ψ(2s) channels to take advantage of the higher statistics of these events.



- To ensure a proper comparison with the MC, we applied the sPlot technique to the collision data, which allows for the statistical subtraction of the background.
- The invariant mass of the final system was used as the discriminating variable in the fit to calculate the statistical weights (sWeights).

$$w_n(x_i) = \frac{\sum_{j=1}^{N_s+N_b} \mathbf{V}_{nj} f_j(x_i)}{\sum_{j=1}^{N_s+N_b} N_j f_j(x_i)}$$

- $f_j(x_i)$ is the PDF value of component *j* for event *i*.
- N_j is the number of events of component j obtained in the fit.
- V_{nj} is the **inverse covariance matrix** of the yields obtained in the likelihood fit.

MC Correction: Reweighting Procedure

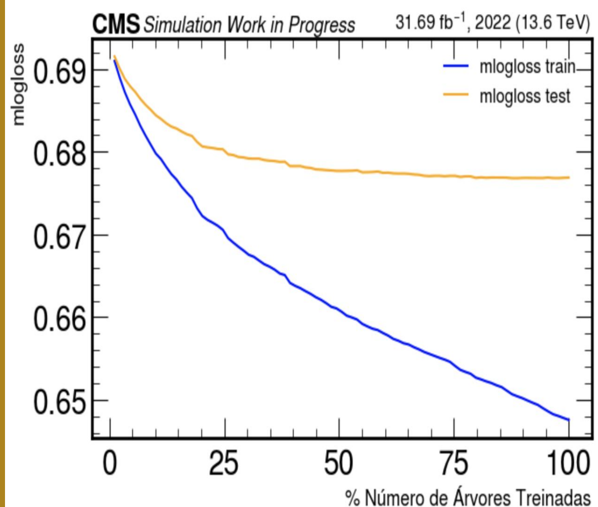
A reweighting technique using the XGBoost Classifier is employed to address discrepancies between data and MC simulation. The classifier is trained to distinguish between data and MC signal distributions, providing event probabilities for both. The ratio of these probabilities approximates the density ratio for reweighting.

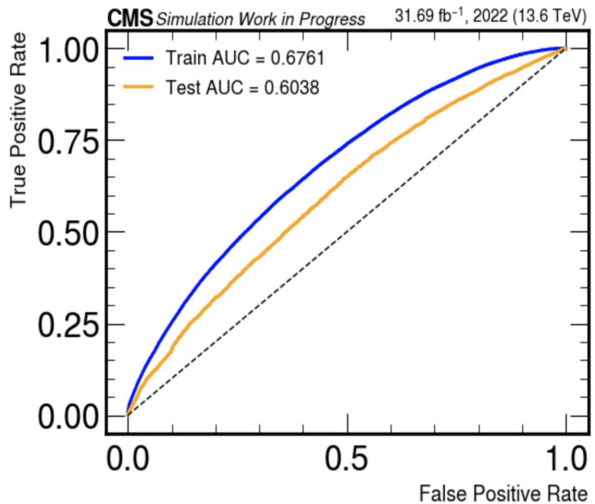
$$w(x) = \frac{f_{data}(x)}{f_{MC}(x)} \sim \frac{p_{data}(x)}{p_{MC}(x)}$$

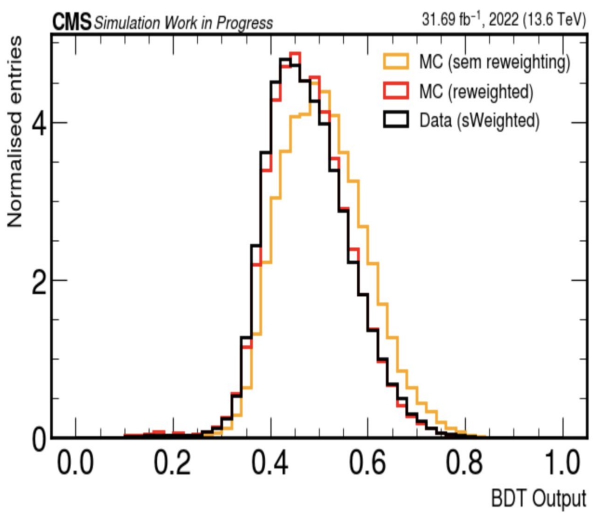
Variable	Description
$p_T(\mu_1)$ [GeV]	p_T of the leading muon
$p_T(\mu_2)$ [GeV]	p_T of the subleading muon
$p_T(trk_1)$ [GeV]	p_T of the leading track
$p_T(trk_2)$ [GeV]	p_T of the subleading track
$\eta(B^0)$	pseudorapidity of the B^0
$\cos\alpha_{2D}$	cosine of the pointing angle between the B^0 momentum and the distance vector from the beamspot to the B^0 vertex, in the transverse plane
DCA sig (trk_1)	transverse impact parameter of the leading track wrt the beamspot
DCA sig (trk_2)	transverse impact parameter of the subleading track wrt the beamspot

Hyperparameter	Value	Description
`objective`	'multi:softprob'	Multiclass probability objective function.
`num_class`	2	Number of target classes.
`eval_metric`	'mlogloss'	Evaluation metric (multiclass log loss).
`learning_rate`	0.1	Learning rate (step size shrinkage) to prevent overfitting.
`max_depth`	7	Maximum depth of the decision trees.
`n_estimators`	500	Number of boosting rounds (trees to build).
`early_stopping_rounds`	10	Validation rounds with no improvement to trigger early stopping.
`random_state`	42	Random seed for reproducibility.
`n_jobs`	-1	Number of parallel threads used to run XGBoost (all available).

MC Correction: Reweighting Procedure

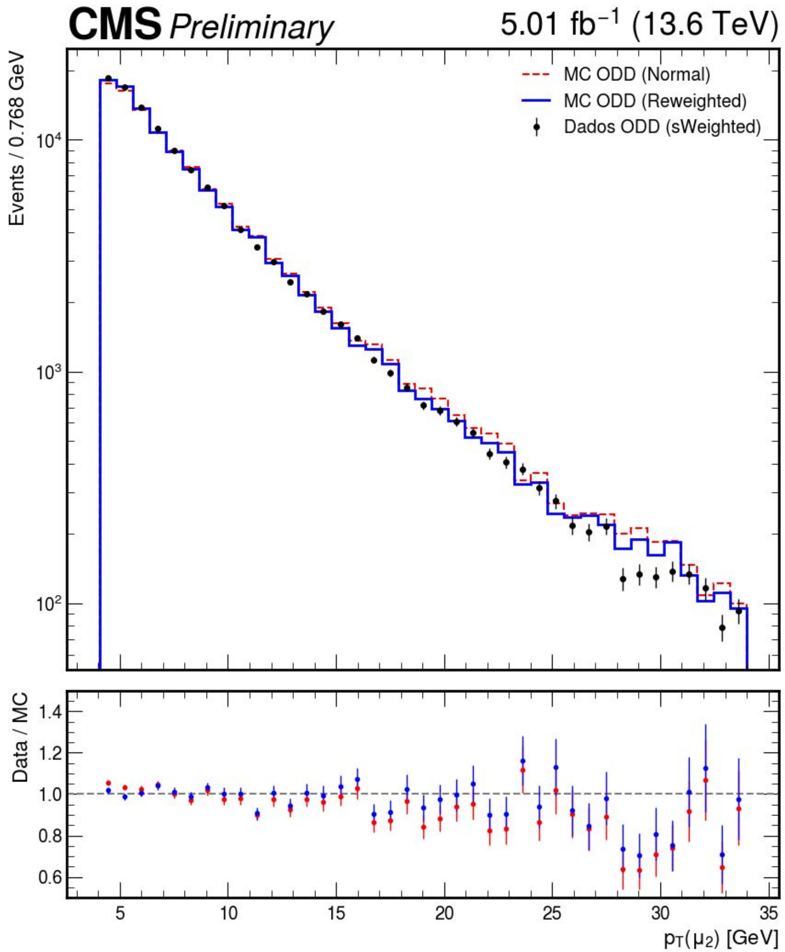


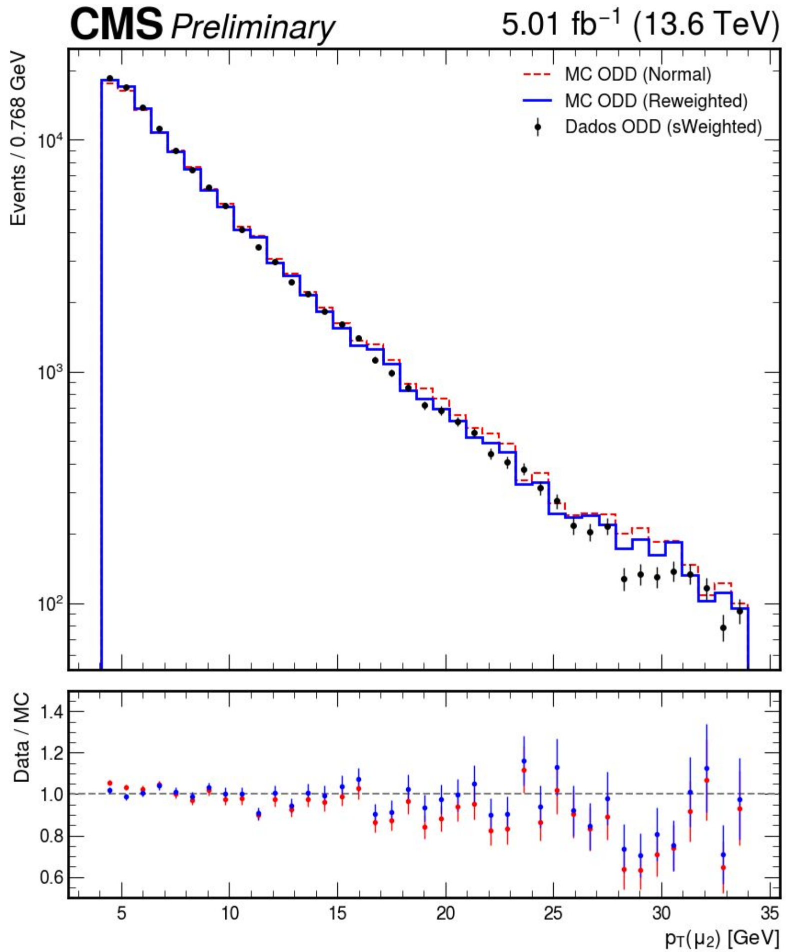




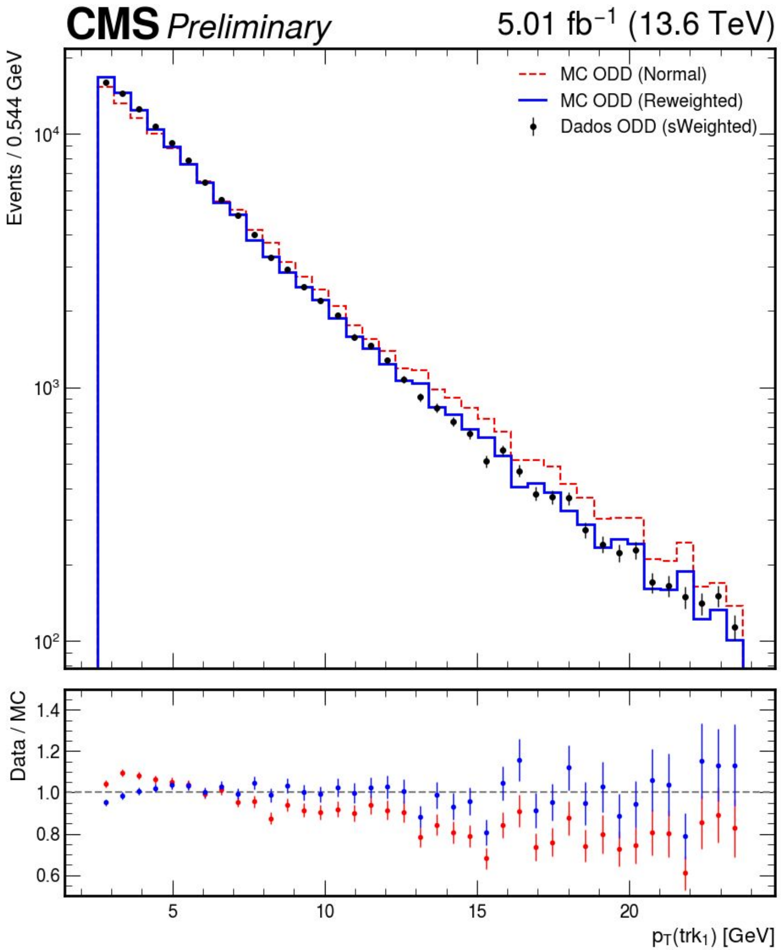
- In these plots, we can see that both the LogLoss and the ROC AUC score exhibit the expected behavior for a successful reweighing procedure.
- The AUC is close to 0.6, and the BDT probability output is heavily centered around 0.4. This indicates that the classifier is effectively guessing at random, but this should be improved to an AUC as close to 0.5 as possible, and a BDT output centered exactly at 0.5, ensuring complete indistinguishability between the Data and the reweighted Monte Carlo.

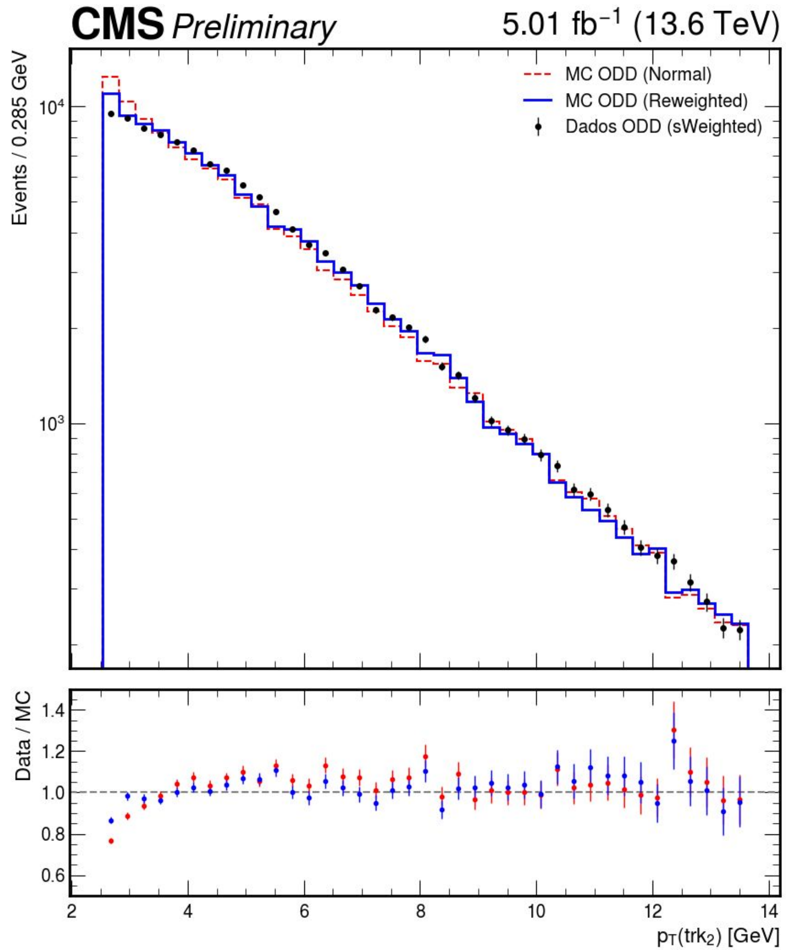
MC Correction: Reweighting Procedure



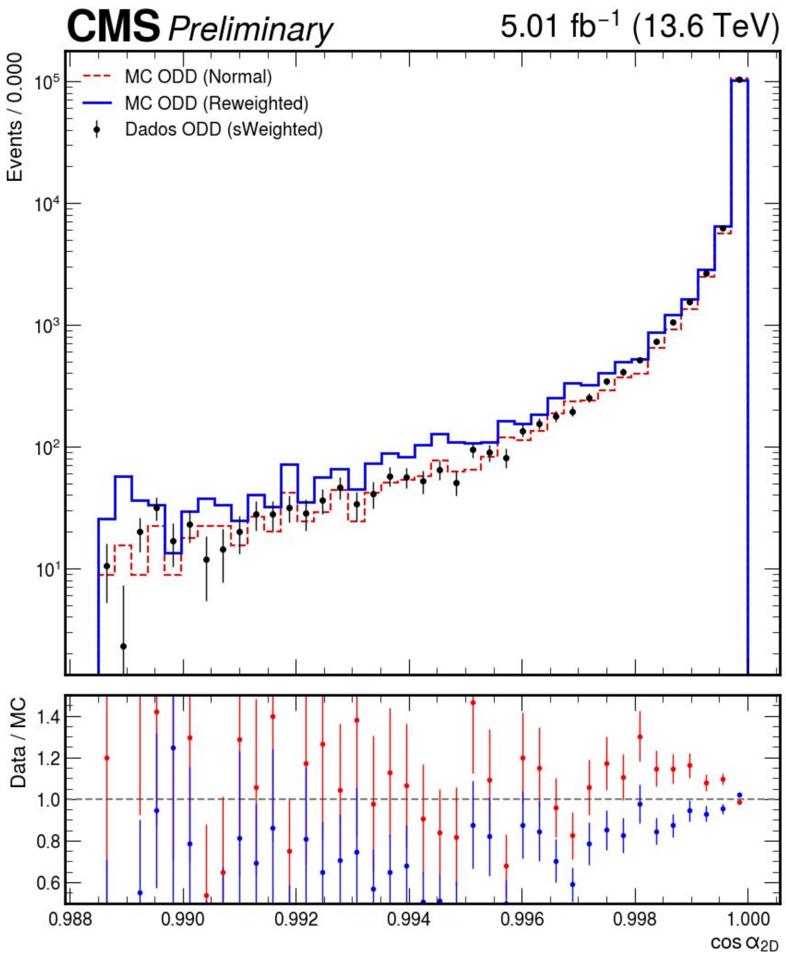


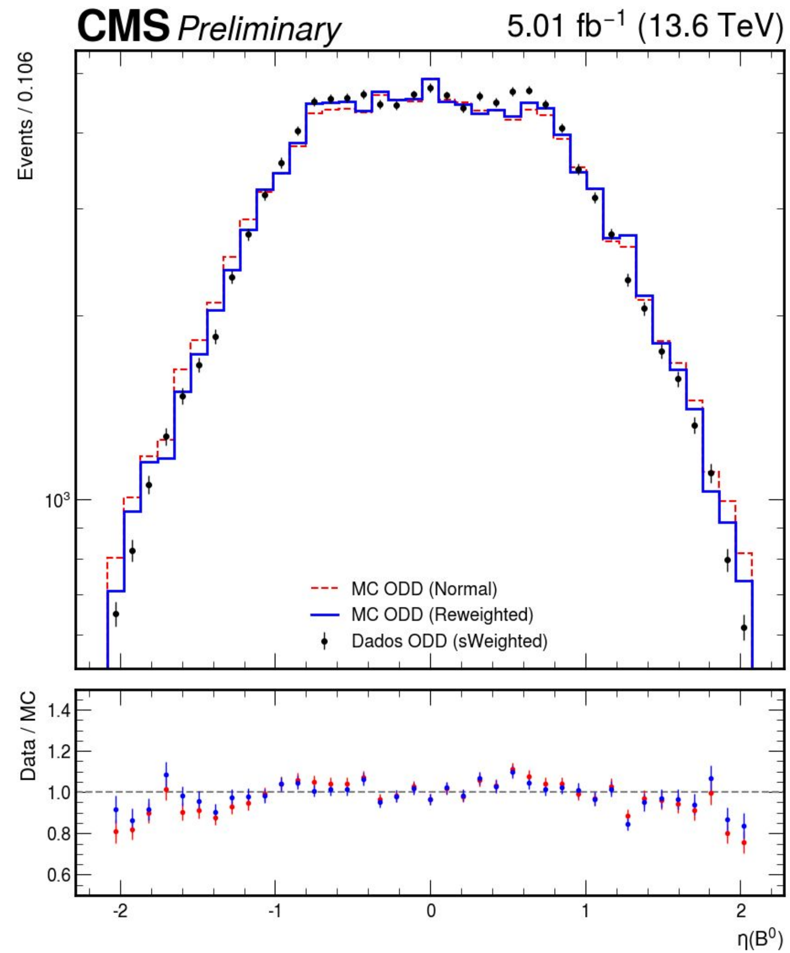
MC Correction: Reweighting Procedure



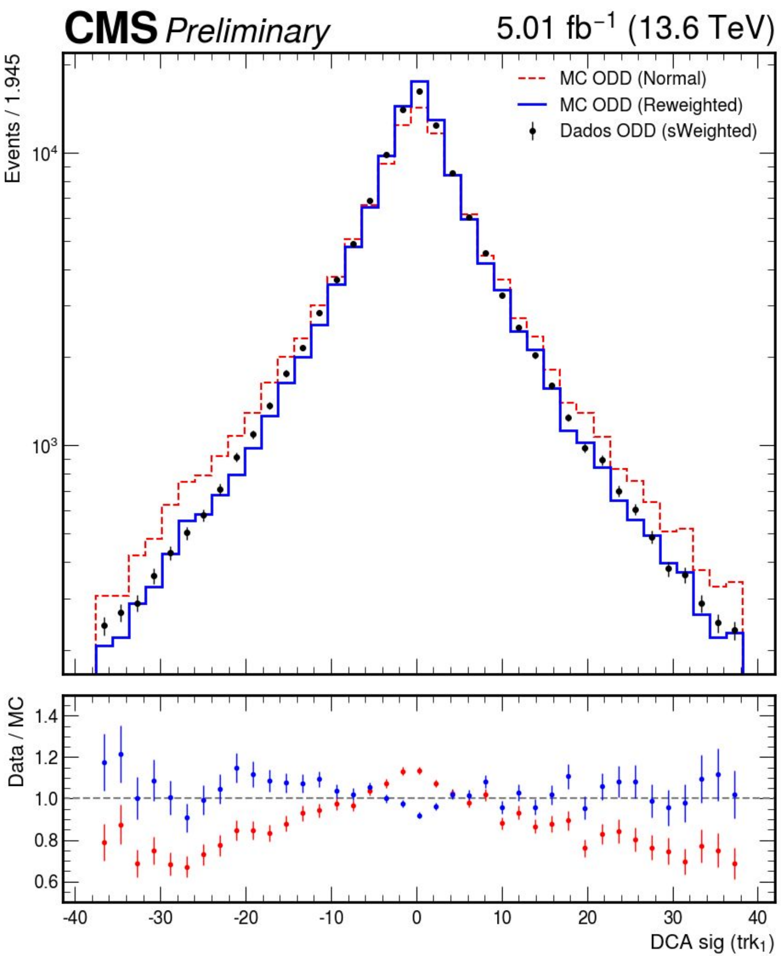


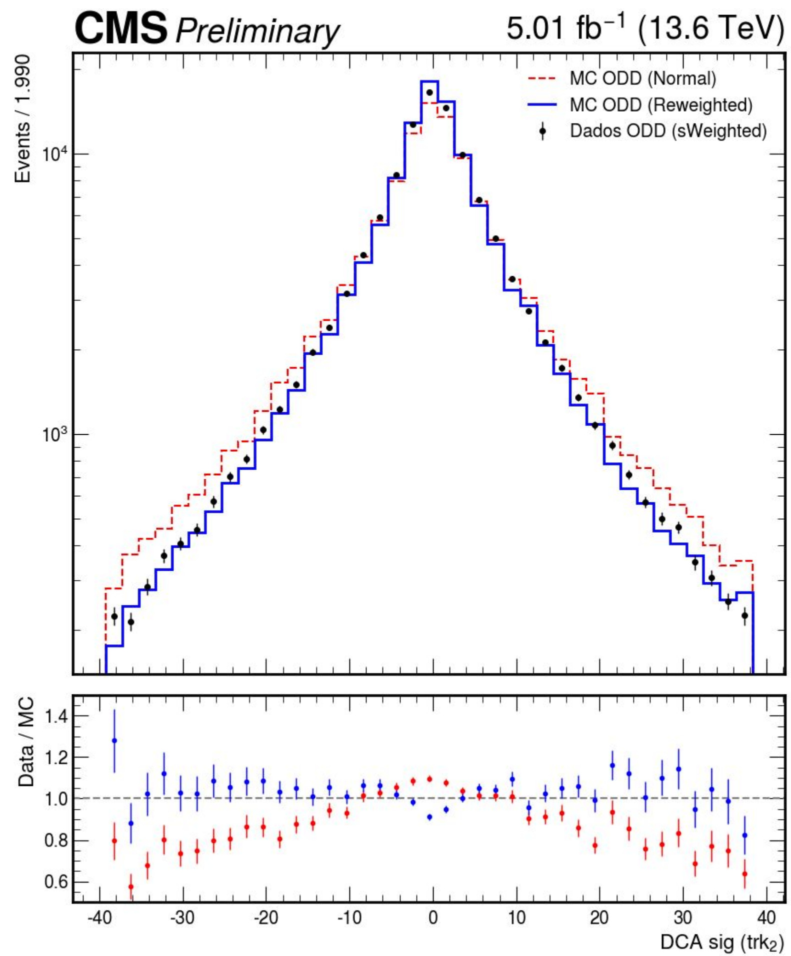
MC Correction: Reweighting Procedure





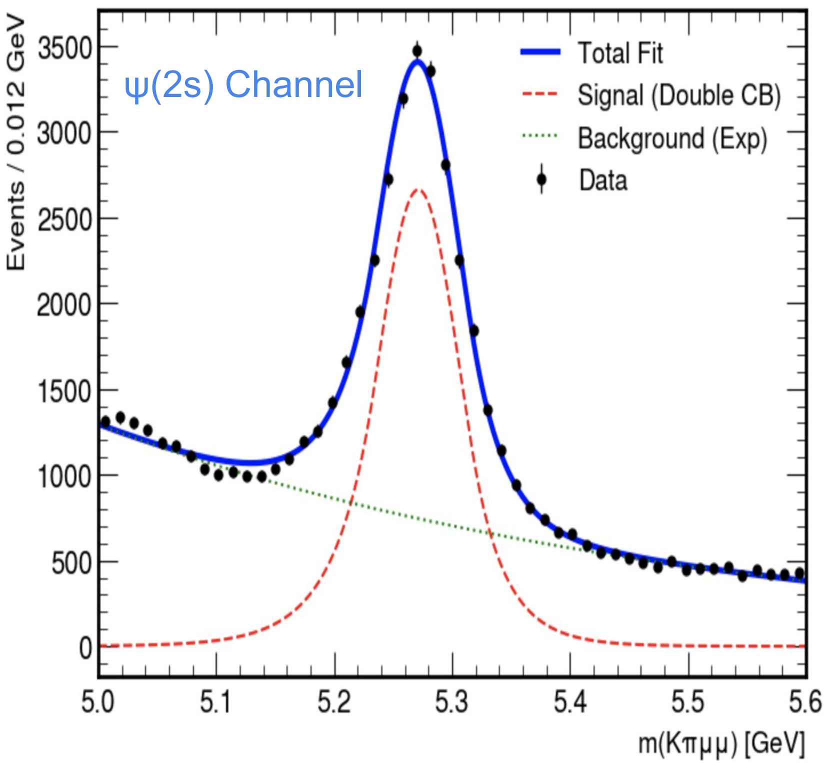
MC Correction: Reweighting Procedure





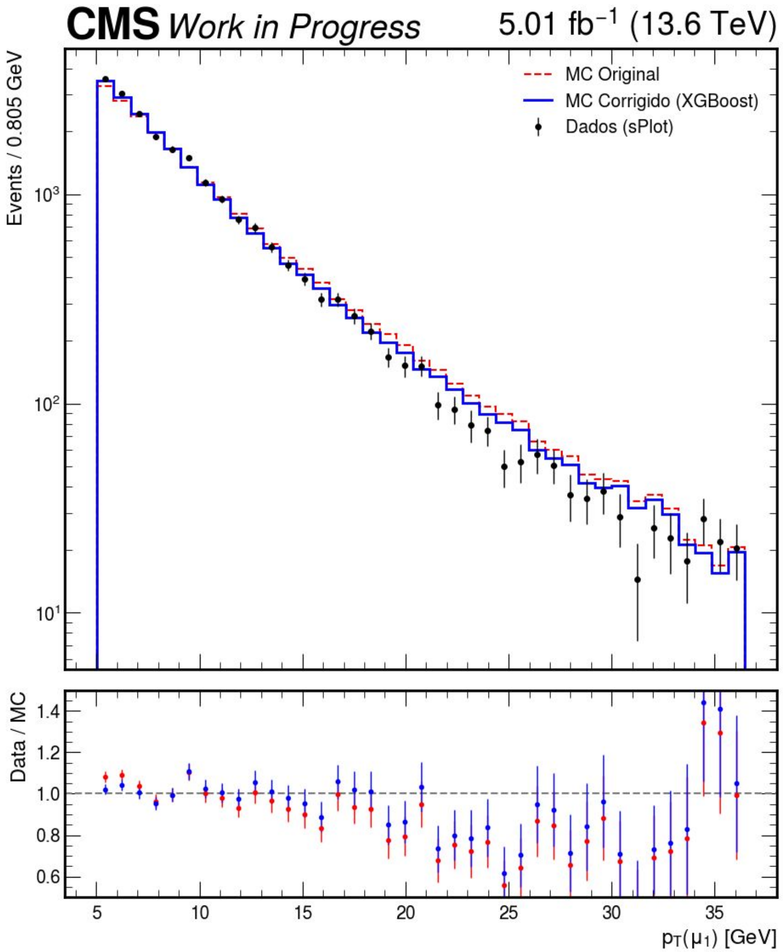
MC Correction: Closure Procedure

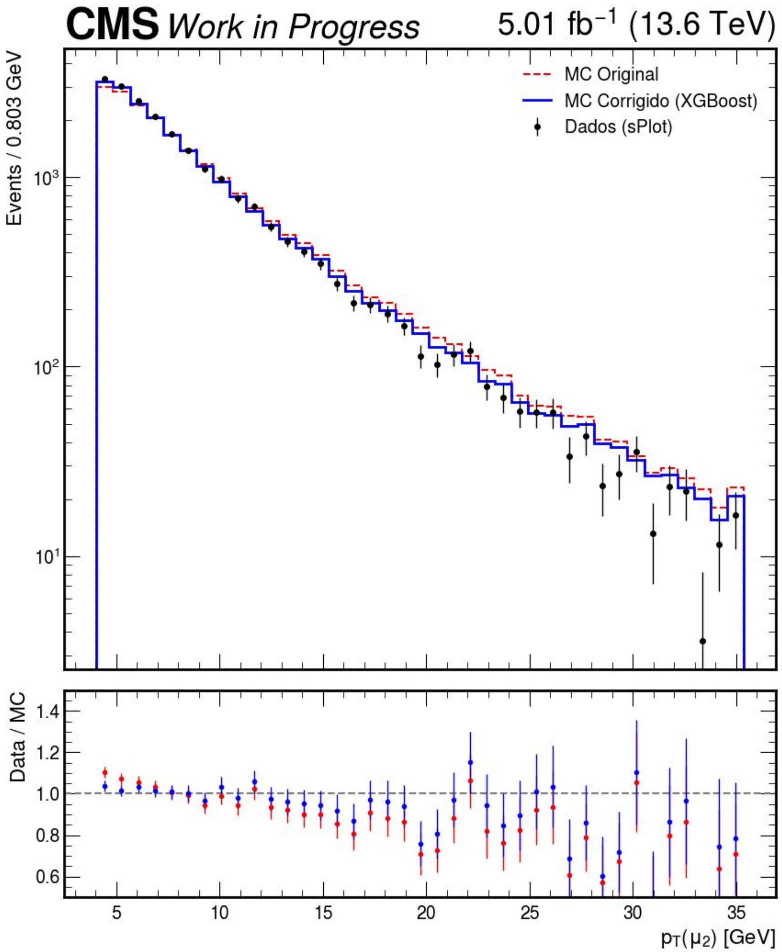
In order to test the model trained in the J/ψ channel we will use this trained model in the ψ(2s) channel as a closure test to check if the model is really correcting the simulation properly.



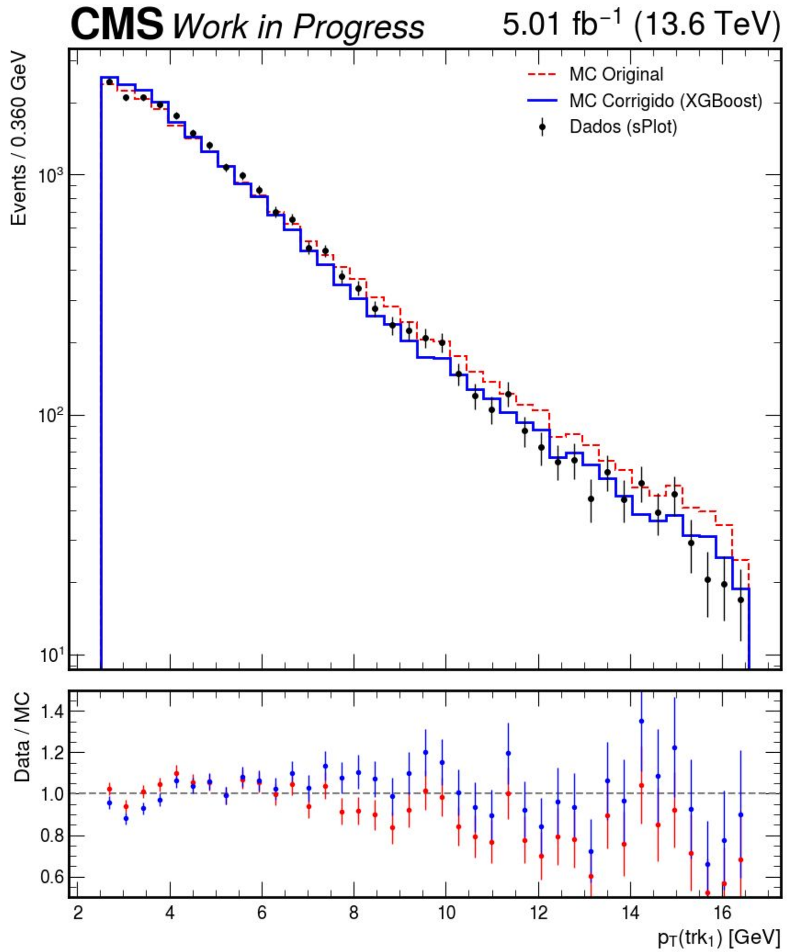
- Application of the weights generated by the BDT (trained on the J/ψ) directly to the B0 → ψ (2S)K*0 control channel sample.
- Use of the sPlot technique on the data to isolate the pure kinematics of the ψ(2S) decay.
- With the weights produced by the BDT model we compare to the data as seen in the next slides:

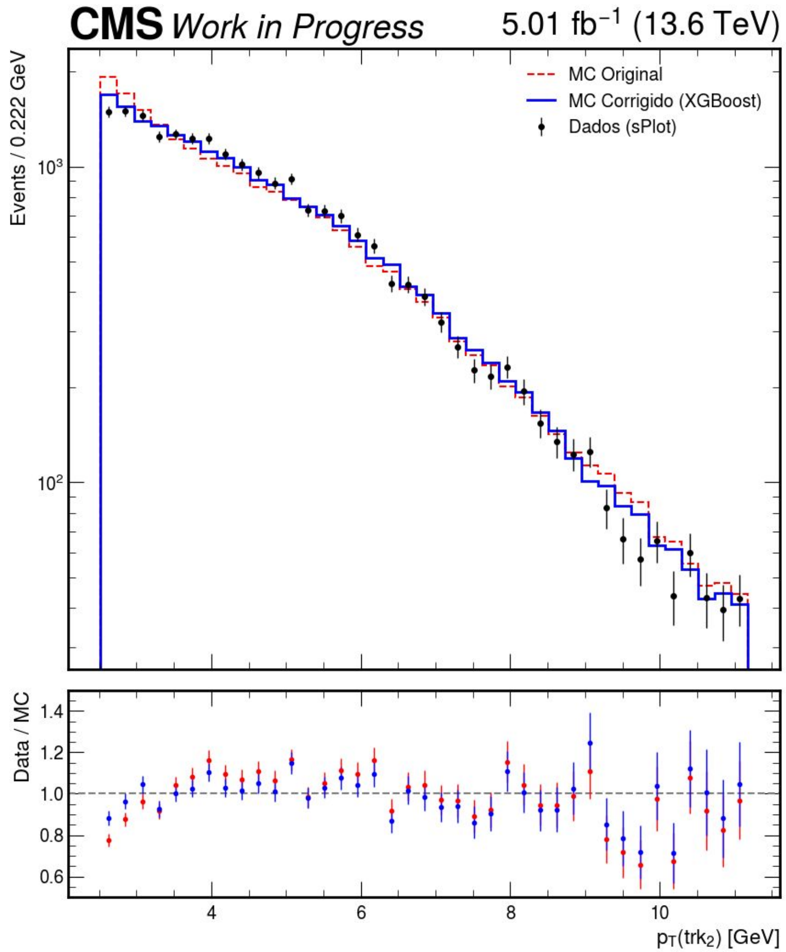
MC Correction: Closure Procedure



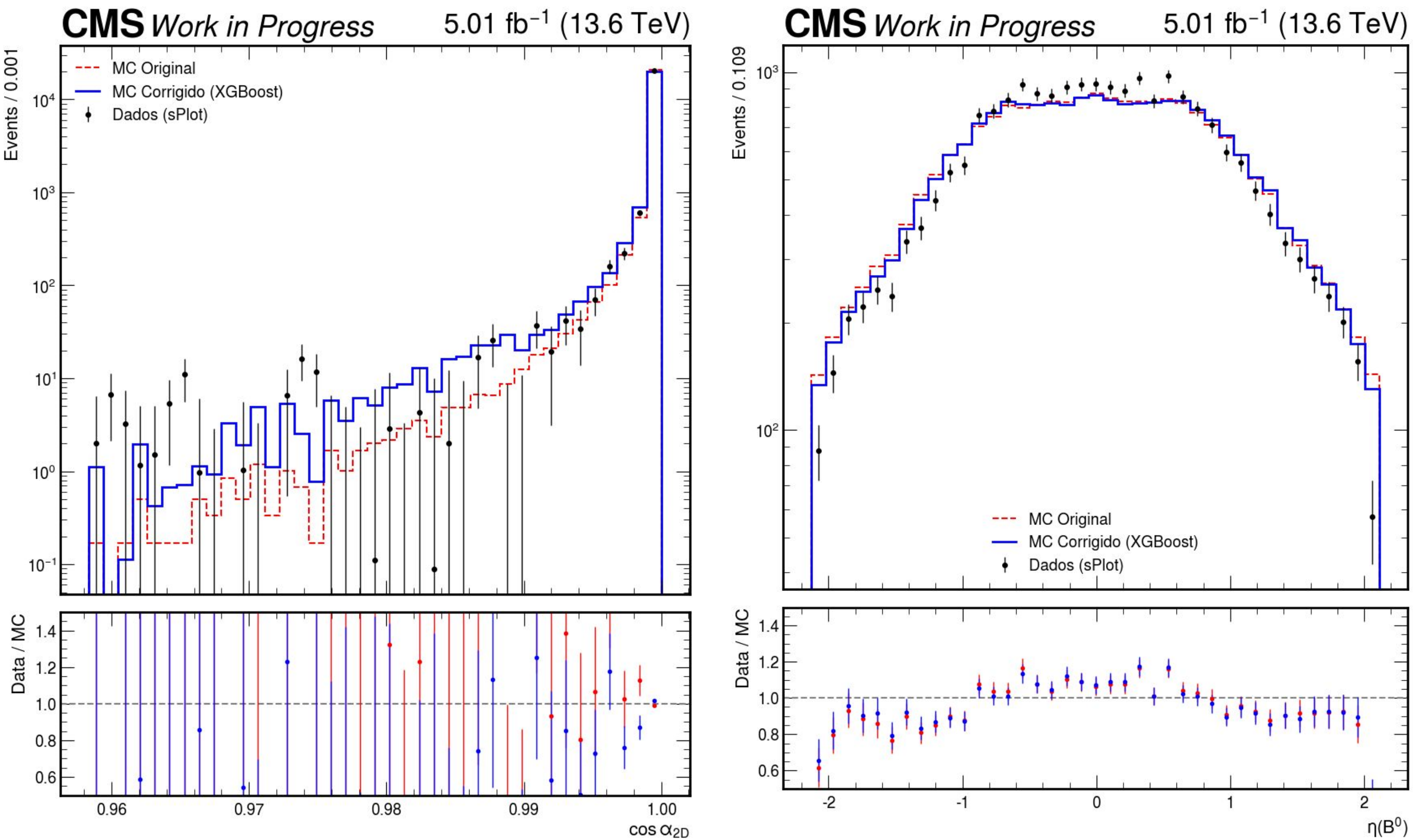


MC Correction: Closure Procedure

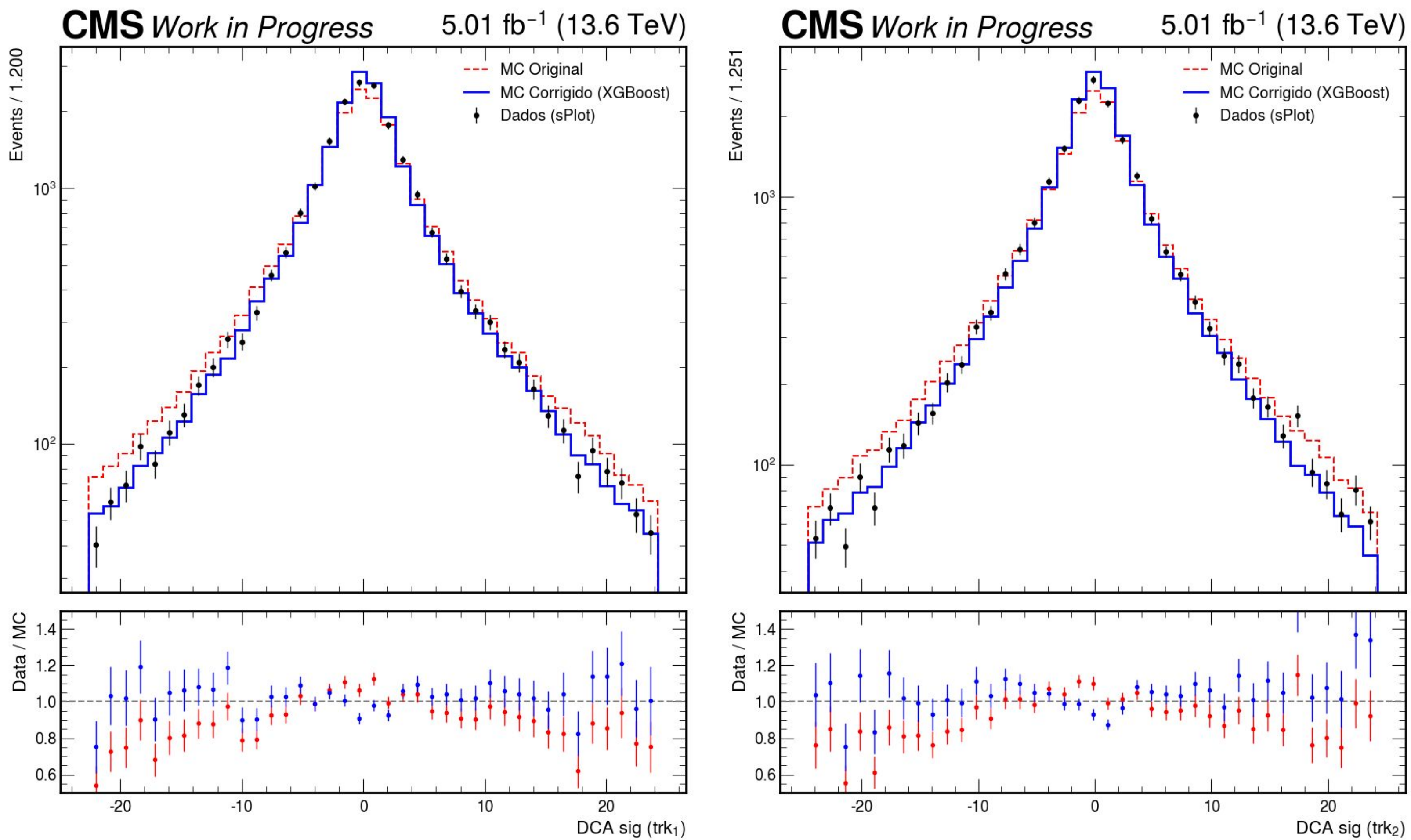




MC Correction: Closure Procedure



MC Correction: Closure Procedure



Next Steps

- **Background Modeling:** Deepen the study and parameterization of physical backgrounds (e.g., peaking backgrounds) that survive the BDT selection.
- **BDT Refinement:** Improve the kinematic reweighting by exploring hyperparameters that further reduce systematic uncertainties.
- **Flavour Mistagging:** Evaluate and correct the meson flavour mistag rate (B^0 vs B^0bar), a critical effect for the measurement of angular asymmetries.
- **Angular Efficiency:** Parameterize the detector acceptance and reconstruction efficiency using the Kernel Density Estimation (KDE) technique.

Thanks for the attention !!!