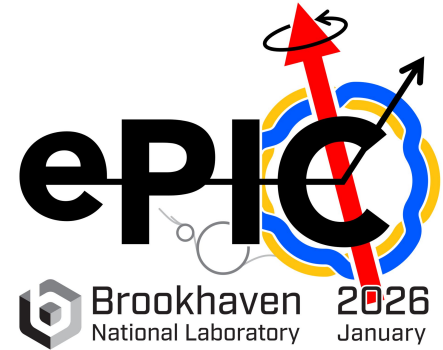


Effects of machine background on D0 invariant mass

Connie Yang, Deepa Thomas
ePIC Collaboration Meeting, Jets and HF PWG Workfest
January 21, 2026



Samples and methods

October ep simulations, D0 and DIS samples, with and without machine background

- 10x100 ep, $Q^2 > 1$, D0 samples \rightarrow signal candidates
- 10x100 ep, $Q^2 > 1$, DIS samples \rightarrow background candidates

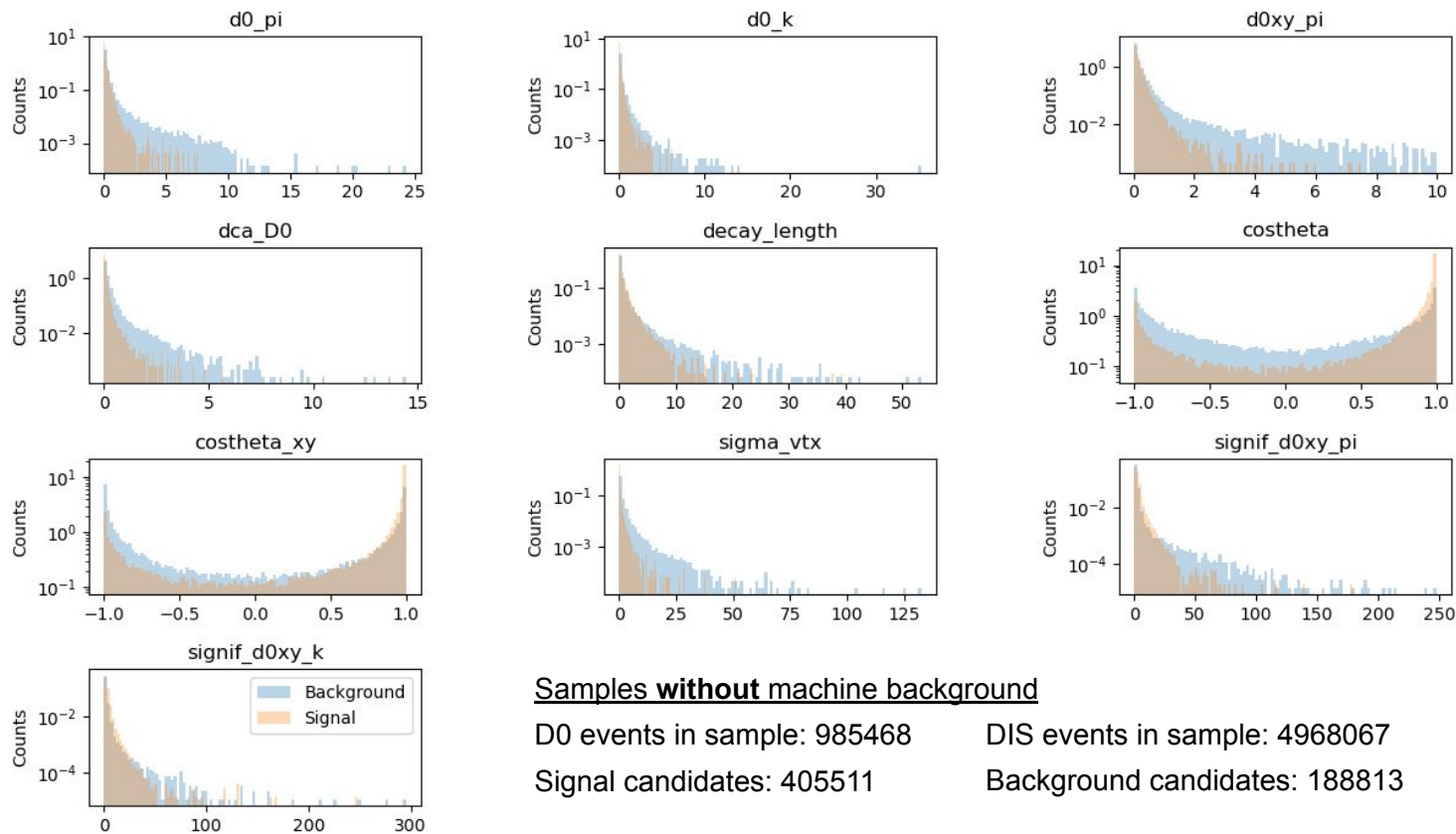
- 10x100 ep, $Q^2 > 1$, D0 + machine bkg samples \rightarrow signal candidates
- 10x100 ep, $Q^2 > 1$, DIS + machine bkg samples \rightarrow background candidates

1 bin with $-2 < y_{D0} < 3$ and $pt_{D0} > 1$

Steps

1. Run analysis to get sig and bkg candidates from samples.
(https://github.com/eic/snippets/tree/main/JetsAndHF/SecondaryVertex_Chi2)
2. Train BDT using sig and bkg candidates, separately for samples with and without machine background.
(https://github.com/eic/snippets/tree/main/JetsAndHF/ML_HF_Reconstruction)
3. Apply topological cuts to sig+bkg candidates using BDT.
4. Compare D0 invariant mass distributions for samples with and without machine bkg.

Without machine bkg, sample statistics and features



Samples **without** machine background

D0 events in sample: 985468

Signal candidates: 405511

DIS events in sample: 4968067

Background candidates: 188813

Preselections

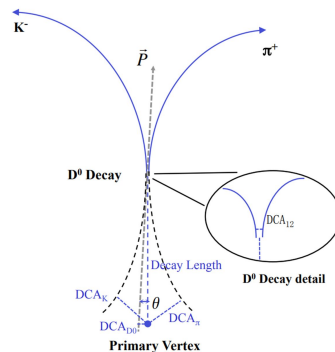
$1.6 < mass_{D0} < 2.5$

$0.02 < d0_{xy_{\pi}} < 10$

$0.02 < d0_{xy_k} < 10$

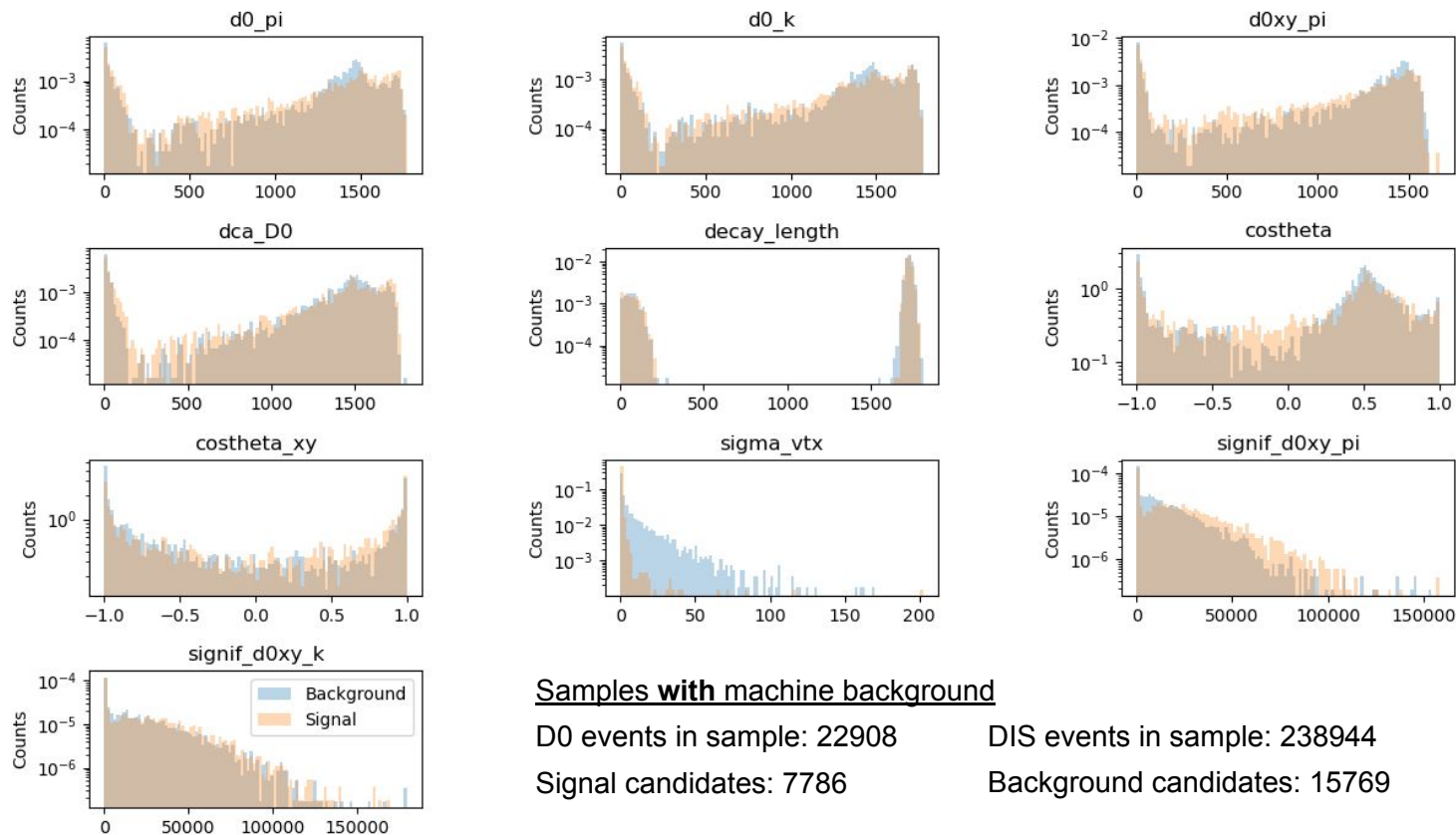
$decay_length < 100$

Features used for training (data after preselections).



Phys. Rev. C 99, 034908 (2019)

With machine bkg, sample statistics and features



Samples with machine background

D0 events in sample: 22908

Signal candidates: 7786

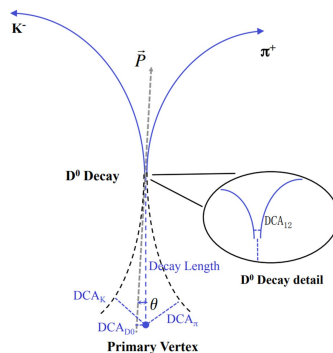
DIS events in sample: 238944

Background candidates: 15769

Preselections

$1.6 < \text{mass}_{D^0} < 2.5$

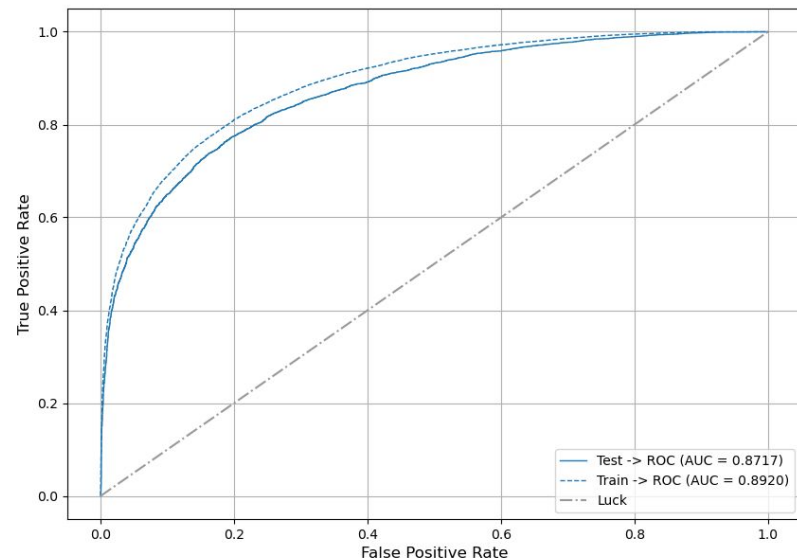
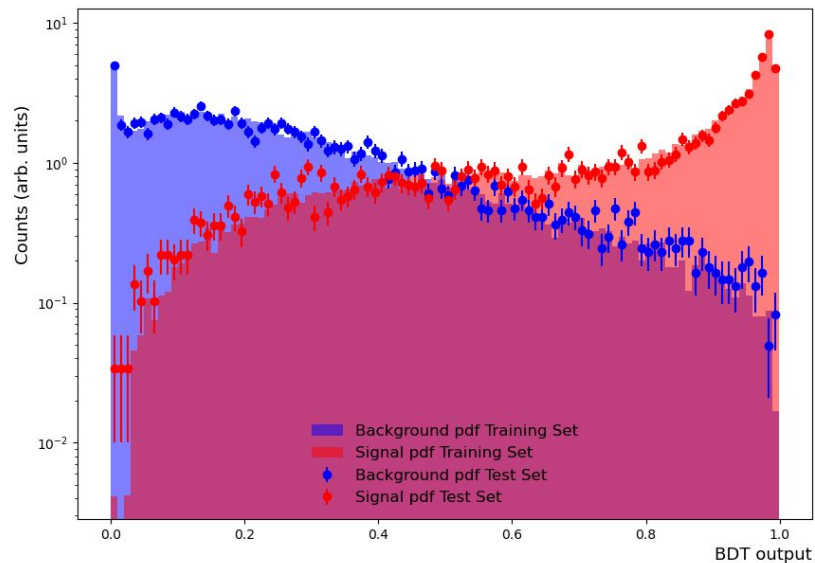
Very large spread in topological distributions with machine bkg.



Phys. Rev. C 99, 034908 (2019)

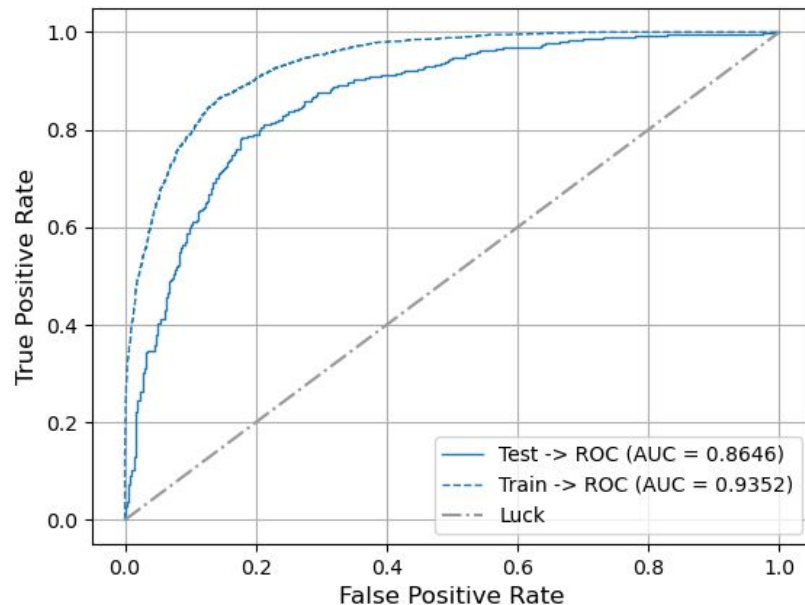
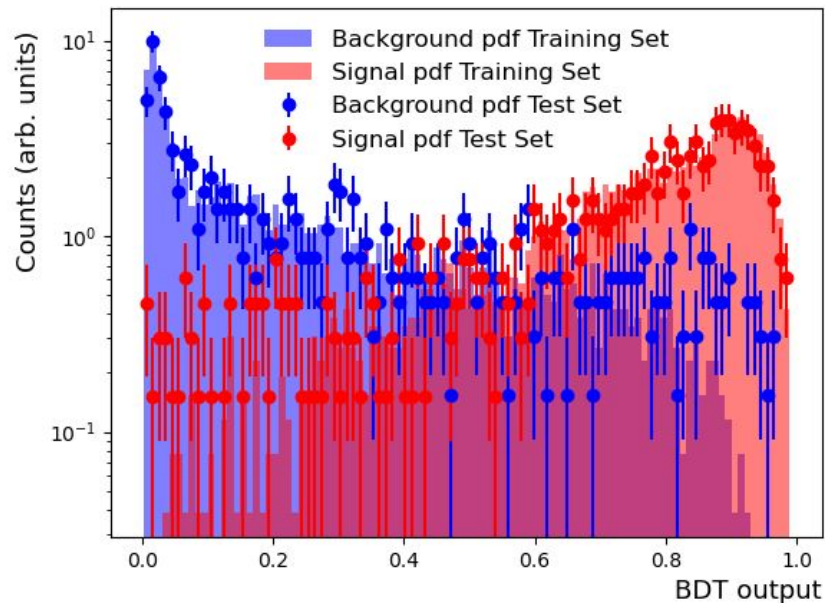
Without machine bkg, BDT results

BDT model is well trained and stable.



With machine bkg, BDT results

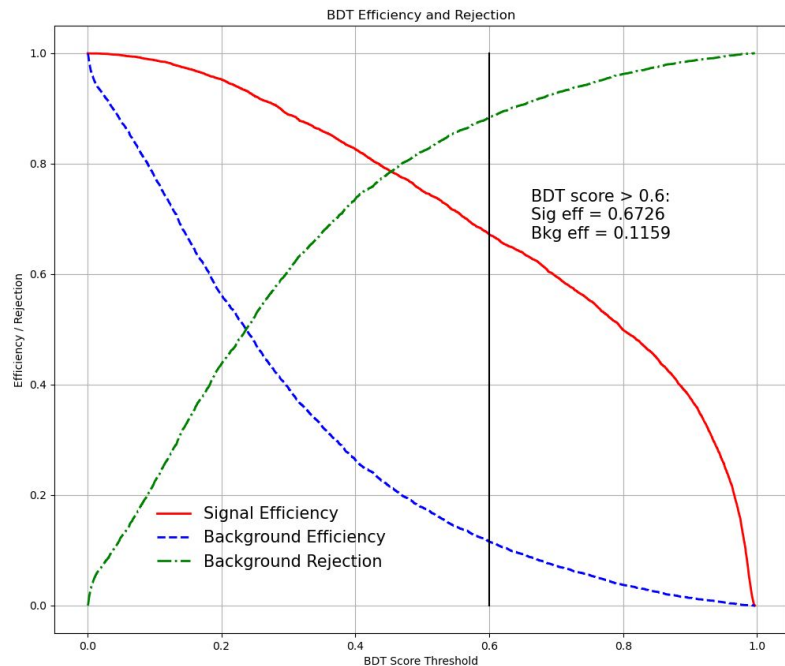
Worse performance for data with machine bkg, BDT training requires further optimization.



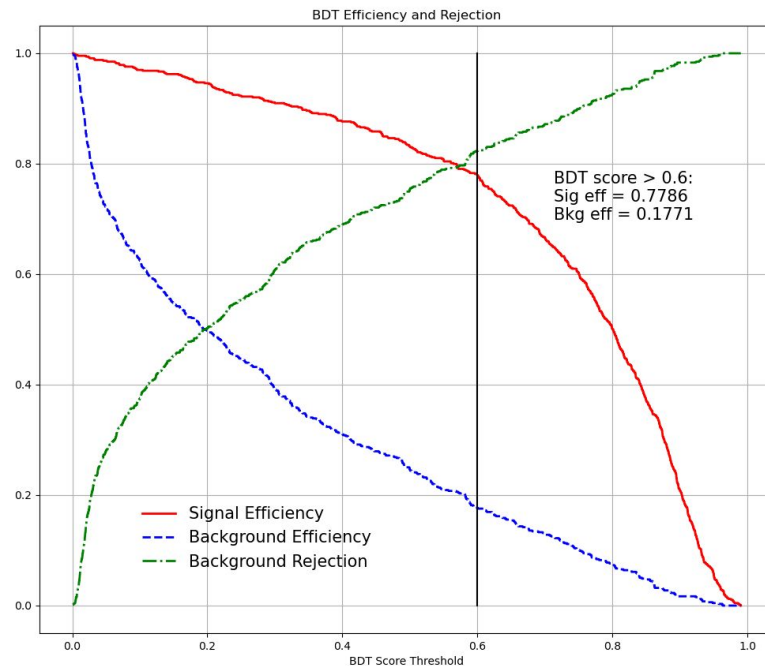
BDT efficiencies

Higher sig eff and bkg eff with machine bkg.

Without machine bkg

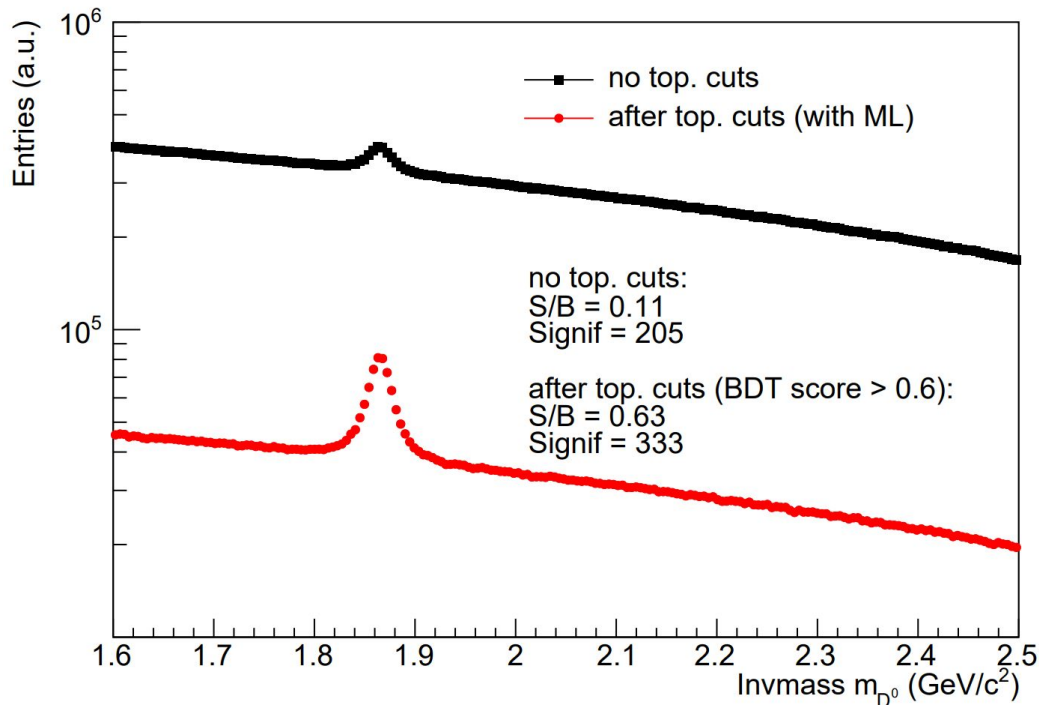


With machine bkg



Without machine bkg, before and after top. cuts

Without machine bkg ($-2 < y < 3$, $pt > 1$)



Nevents D0:

$$985468 * 1821.83 \sim 1.80 \times 10^9$$

Nevents DIS: 4968067

$$N_{\text{resampled}} = \frac{N_{\text{candidates}}}{N_{\text{events}}} \times 4.72162e9$$

Expected Nevents
for 10 fb⁻¹

$$N_{\text{resampled}}^{\text{ML}} = N_{\text{resampled}} \times \text{eff}(\text{BDT} = 0.6)$$

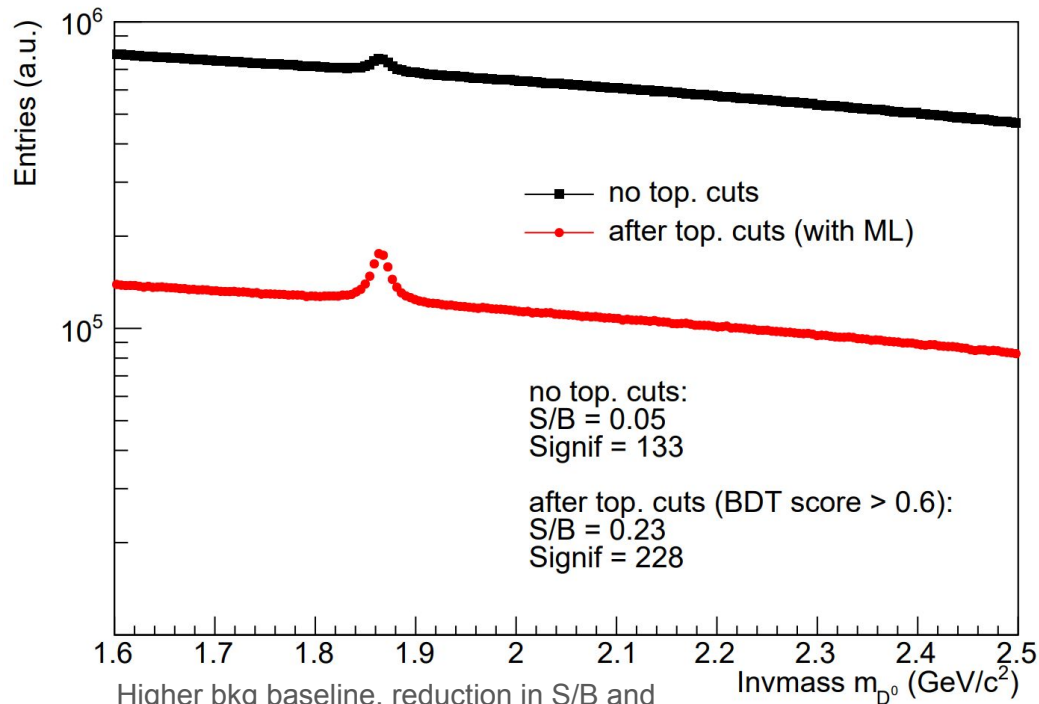
Signal peak: $1.84 < \text{Invmass} < 1.89$

$$S/B = N_{\text{resampled}}^{\text{sig}} / N_{\text{resampled}}^{\text{bkg}}$$

$$\text{Signif} = N_{\text{resampled}}^{\text{sig}} / \sqrt{N_{\text{resampled}}^{\text{sig}} + N_{\text{resampled}}^{\text{bkg}}}$$

With machine bkg, before and after top. cuts

With machine bkg ($-2 < y < 3$, $pt > 1$)



Higher bkg baseline, reduction in S/B and Signif with machine bkg.

Nevents D0:

$$22908.0 * 1821.83 \sim 4.17 \times 10^7$$

Nevents DIS: 238944

$$N_{\text{resampled}} = \frac{N_{\text{candidates}}}{N_{\text{events}}} \times 4.72162e9$$

Expected Nevents
for 10 fb^{-1}

$$N_{\text{resampled}}^{\text{ML}} = N_{\text{resampled}} \times \text{eff}(\text{BDT} = 0.6)$$

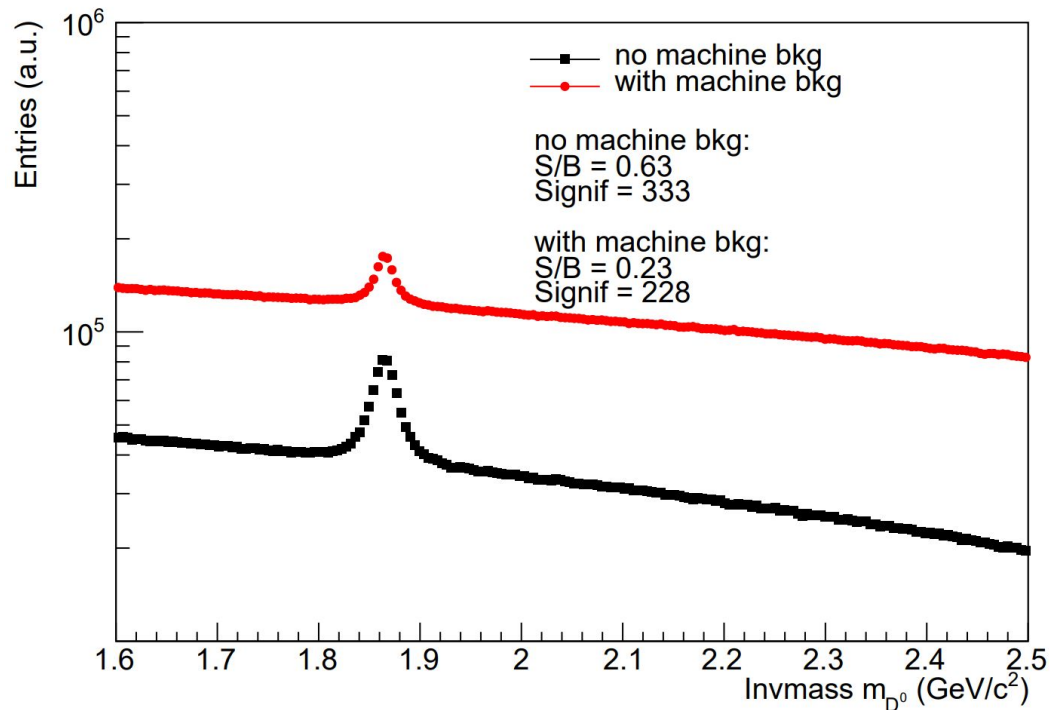
Signal peak: $1.84 < \text{Invmass} < 1.89$

$$S/B = N_{\text{resampled}}^{\text{sig}} / N_{\text{resampled}}^{\text{bkg}}$$

$$\text{Signif} = N_{\text{resampled}}^{\text{sig}} / \sqrt{N_{\text{resampled}}^{\text{sig}} + N_{\text{resampled}}^{\text{bkg}}}$$

Summary and outlook

After top. cuts (BDT score > 0.6, $-2 < y < 3$, $p_t > 1$)



- Trained BDT model on D0 and DIS samples.
- Applied topological cuts.
- Obtained D0 invariant mass distributions with and without machine bkg.
- Observed reduction in S/B and Signif with machine bkg.

Next steps

- Optimize BDT training.
- Examine primary vertex distribution for samples with machine bkg.

References and acknowledgements

October ep simulations

With machine background

D0: /volatile/eic/EPIC/RECO/25.10.4/epic_craterlake/Bkg_Exactly1SignalPer2usFrame/SIDIS/D0_ABCONV/pythia8.306-1.1/10x100/q2_1/hiDiv/

DIS: /volatile/eic/EPIC/RECO/25.10.4/epic_craterlake/Bkg_1SignalPer2usFrame/DIS/NC/10x100/minQ2=1

Without machine background

D0: /volatile/eic/EPIC/RECO/25.10.3/epic_craterlake/SIDIS/D0_ABCONV/pythia8.306-1.1/10x100/q2_1/hiDiv/

DIS: /volatile/eic/EPIC/RECO/25.10.0/epic_craterlake/DIS/NC/10x100/minQ2=1

Analysis framework for sig and bkg candidates: https://github.com/eic/snippets/tree/main/JetsAndHF/SecondaryVertex_Chi2

Machine learning framework: https://github.com/eic/snippets/tree/main/JetsAndHF/ML_HF_Reconstruction

Based on hipe4ml: <https://doi.org/10.5281/zenodo.5070131>

Thanks to Deepa Thomas, Shyam Kumar, Rongrong Ma

Thank you