VBS VVH All-Hadronic Status Report

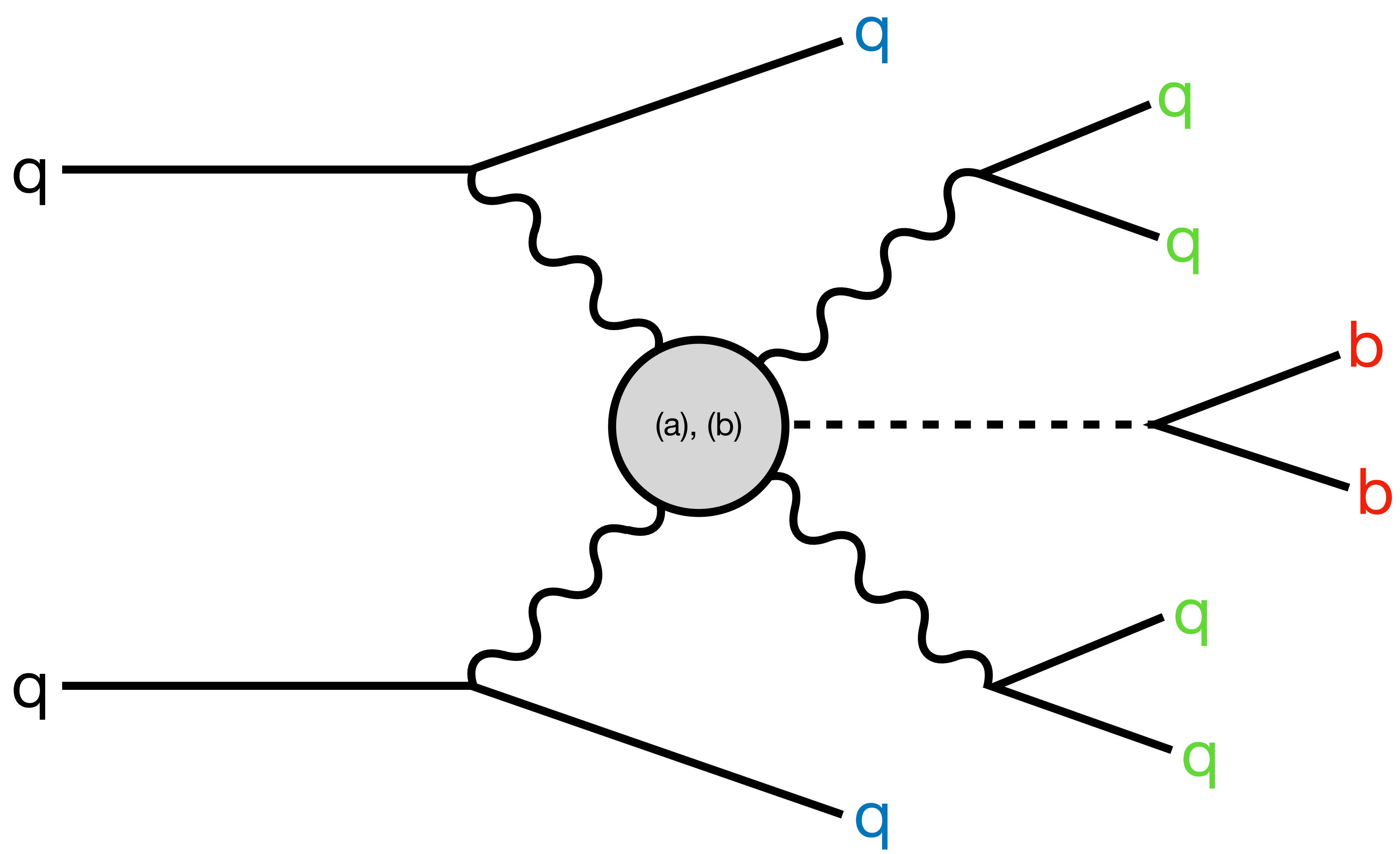
A. Arora, R. Band, P. Chang, L. Giannini, **J. Guiang**, C. Jessop, L. Lutton,
F. Würthwein, Y. Xiang, A. Yagil, Y. Zhou, L. Zygala

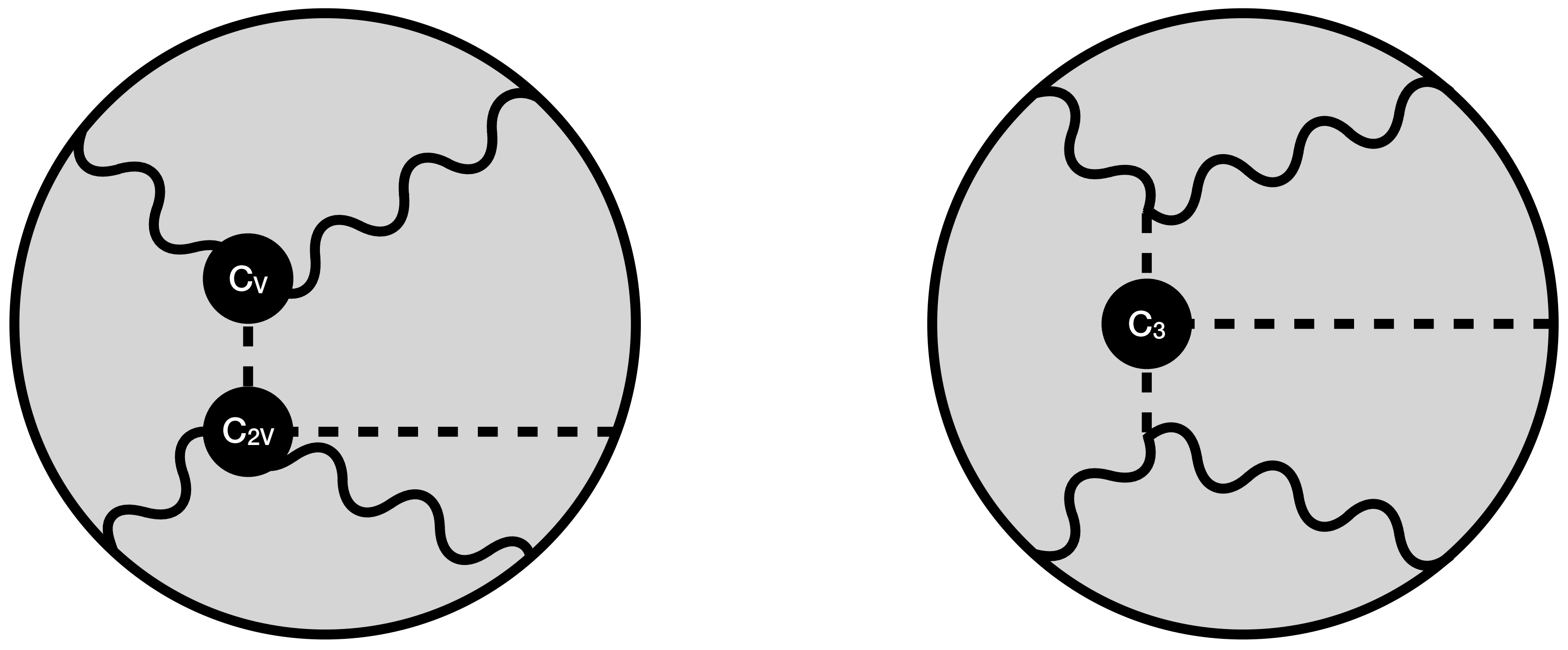
July 26th, 2023

Analysis Overview

All-Hadronic VBS VVH

- Targeting VBS WWH, WZH, ZZH in the all-hadronic final state
- Sensitive to C_{2V}, C_3, and C_V in principle
    - **We focus on C_{2V} for now (most sensitive)**
- Signal signature:
    - W/Z/H jets with large p_T for e.g. BSM C_{2V}
    - VBS jets with large $\Delta\eta_{jj}$, M_{jj}
- **This analysis is optimized for $C_{2V} = 2$**



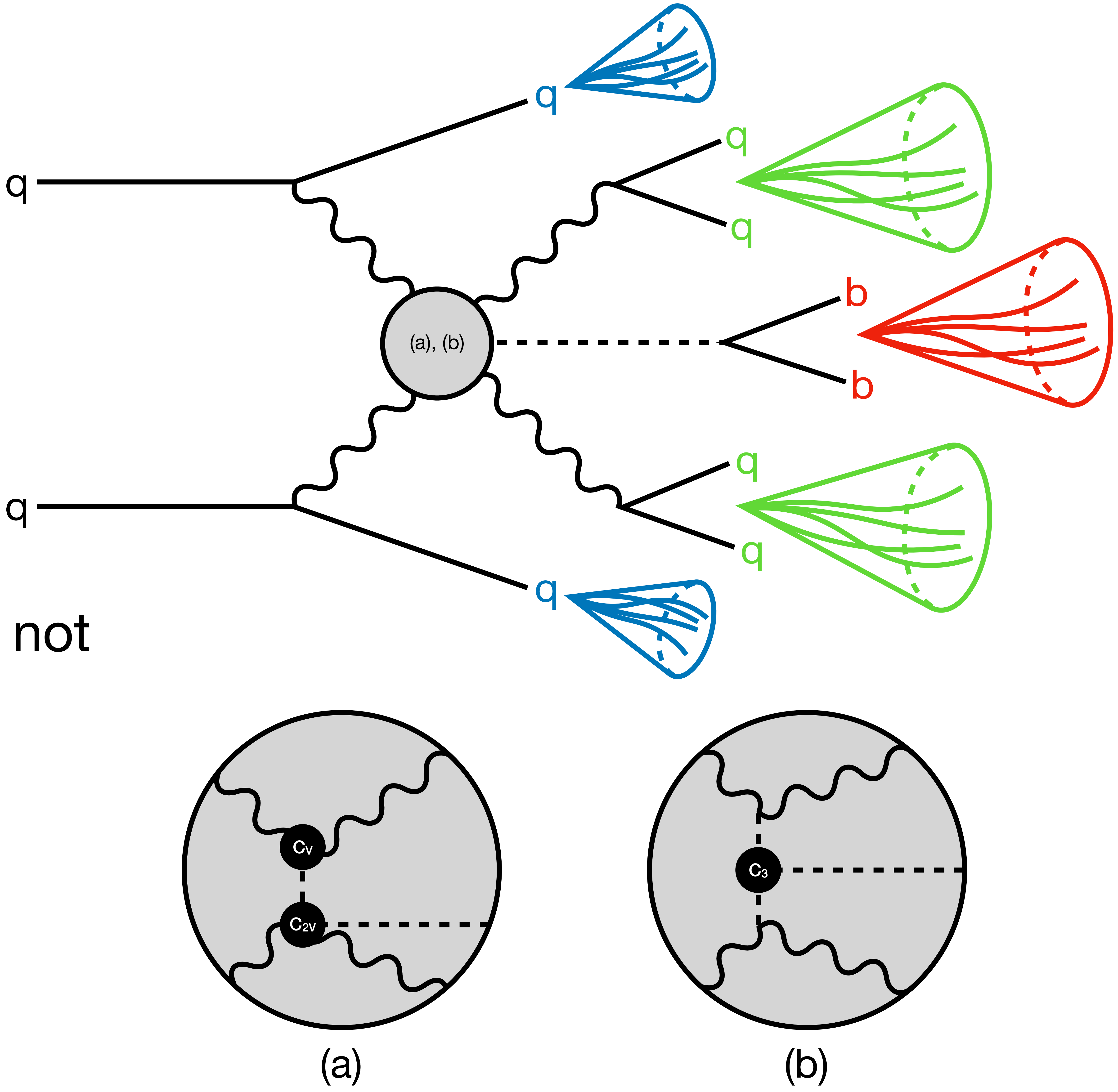


(a) (b)

UC San Diego

All-Hadronic VBS VVH

- One interesting N_{jets} vs. $N_{fatjets}$ channel:
  - **≥ 3 AK8 fat jets, ≥ 2 AK4 jets** (right)
  - ~~2 AK8 fat jets, ≥ 4 AK4 jets~~
  - ~~2 AK8 fat jets, 3 AK4 jets~~
- From previous studies, $N_{fatjets} < 3$ channels not worthwhile pursuing right now
- **We focus only on the 3 fat jet channel**



Skim + Triggers + 3 Fat Jet Region

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}$+1ℓ	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%

Object	Skim Selection
Leptons (μ, e)	== 0 veto*
Fat Jets	**≥ 2 AK8 jets w/ p_T > 300 GeV** AND $\lvert\eta\rvert$ < 2.5 AND mass > 50 GeV AND M_{SD} > 40 GeV AND fat jet ID > 0 AND max(p_T) > 550 GeV**
Jets	≥ 2 AK4 jets w/ p_T > 30 GeV AND passes tight jet ID AND ΔR(jet, fat jet) > 0.8

*Using the ttH lepton ID
**HLT plateau

Year	HLT path
2016	HLT_PFHT800 HLT_PFHT900 HLT_AK8PFHT650_TrimR0p1PT0p03Mass50 HLT_AK8PFHT700_TrimR0p1PT0p03Mass50 HLT_AK8PFJet450 HLT_AK8PFJet360_TrimMass30 HLT_AK8DiPFJet280_200_TrimMass30 HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV_p20
2017	HLT_PFHT1050 HLT_AK8PFHT800_TrimMass50 HLT_PFJet320 HLT_PFJet500 HLT_AK8PFJet320 HLT_AK8PFJet500 HLT_AK8PFJet400_TrimMass30 HLT_AK8PFJet420_TrimMass30
2018	HLT_PFHT1050 HLT_AK8PFHT800_TrimMass50 HLT_PFJet500 HLT_AK8PFJet500 HLT_AK8PFJet400_TrimMass30 HLT_AK8PFJet420_TrimMass30

Taken from B2G-21-003

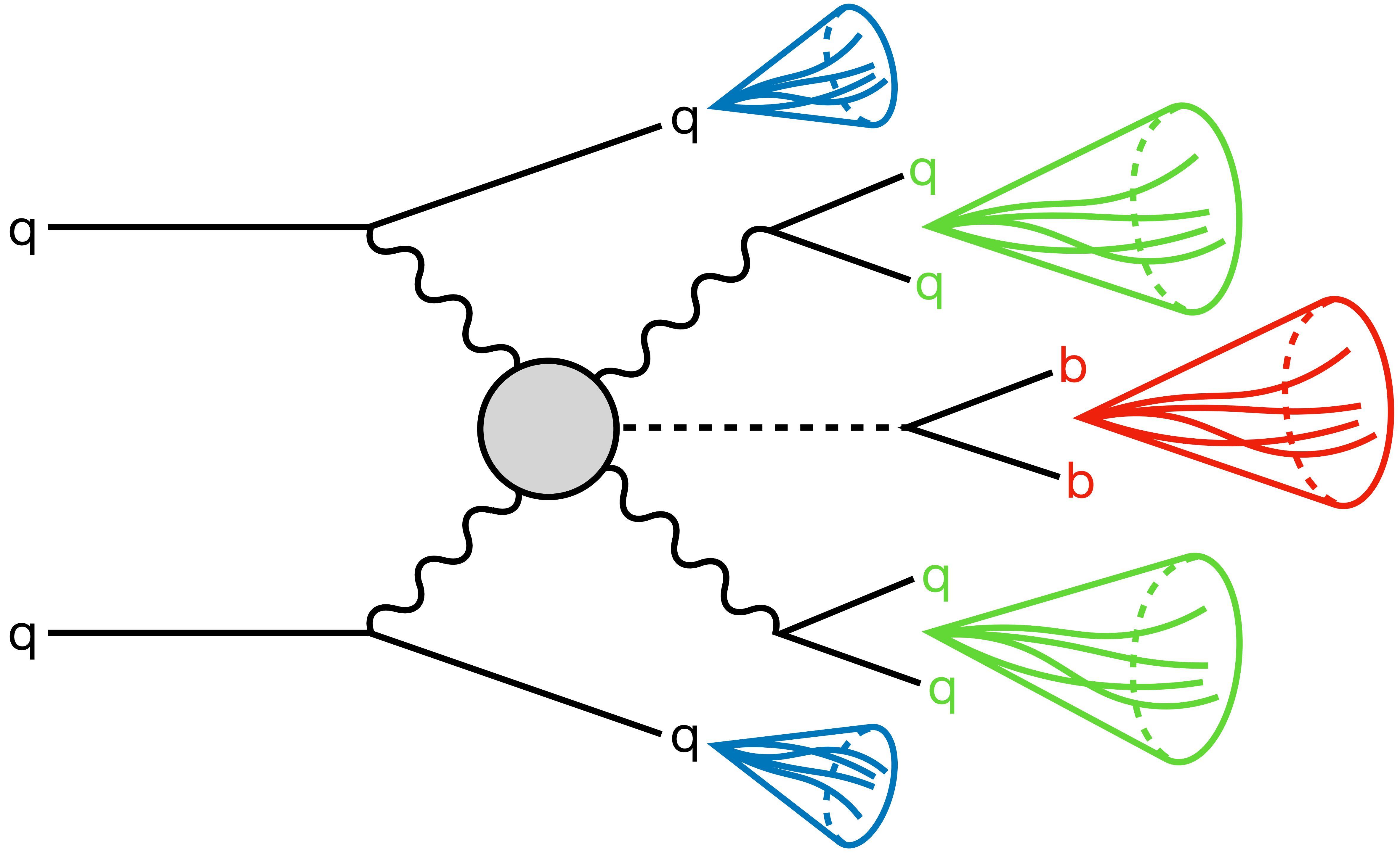
UC San Diego

Object Selection

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}+1\ell$	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%
Object selection	158K	6.2K	855	59	30	478	5.1K	171K	41%	18	56%

Object	Selections
H→bb fat jet	• Has max(ParticleNet Xbb)
V→qq fat jets	• Not the H→bb candidate • Leading and next-leading in p_T
VBS (AK4) jets	• ΔR(jet, H→bb fat jet) > 0.8 • ΔR(jet, lead V→qq fat jet) > 0.8 • ΔR(jet, trail V→qq fat jet) > 0.8 • For > 2 candidates: • Take pair with maximum $\lvert\Delta\eta_{jj}\rvert$



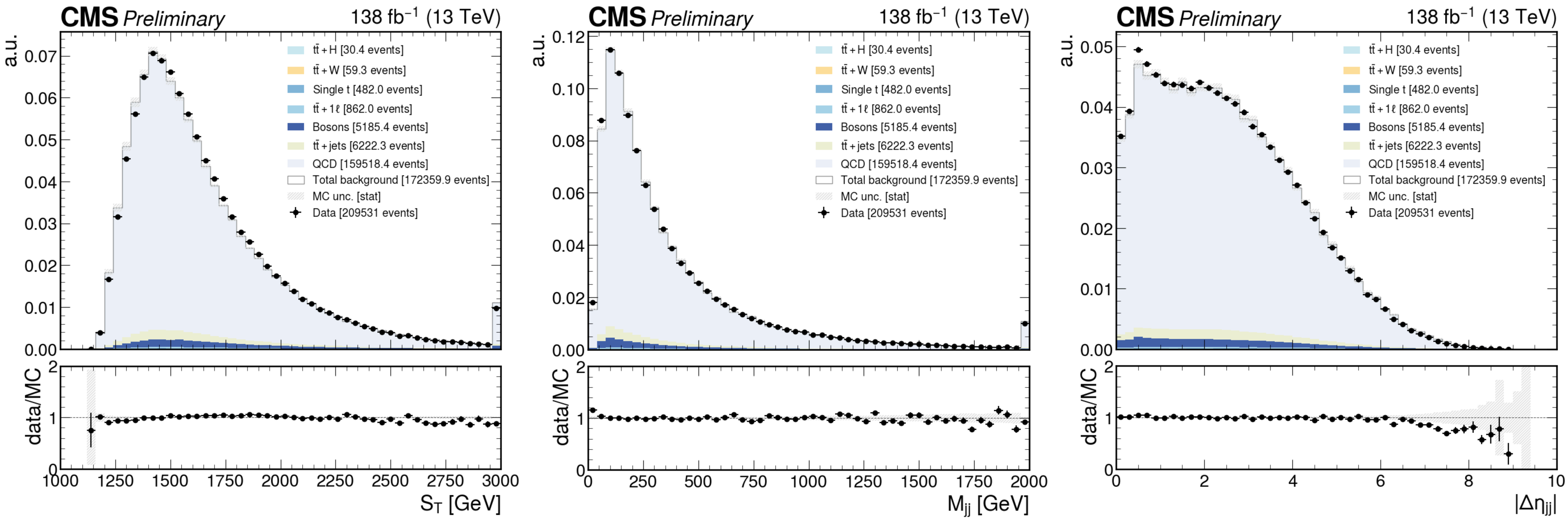
Object Selection

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}$+1ℓ	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%
Object selection	158K	6.2K	855	59	30	478	5.1K	171K	41%	18	56%

- **Next:** plot data/MC here and check for agreement
- We ultimately estimate bkg. From MC, so only a minor check:
    - Looking for any systematic disagreement
- Safe to unblind at Object Selection since signal is so small

Data vs. MC



Data and MC agree well for most variables (more plots here)

UC San Diego

Preselection

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}+1\ell$	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH ($C_{2V}=2$)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%
Object selection	158K	6.2K	855	59	30	478	5.1K	171K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%

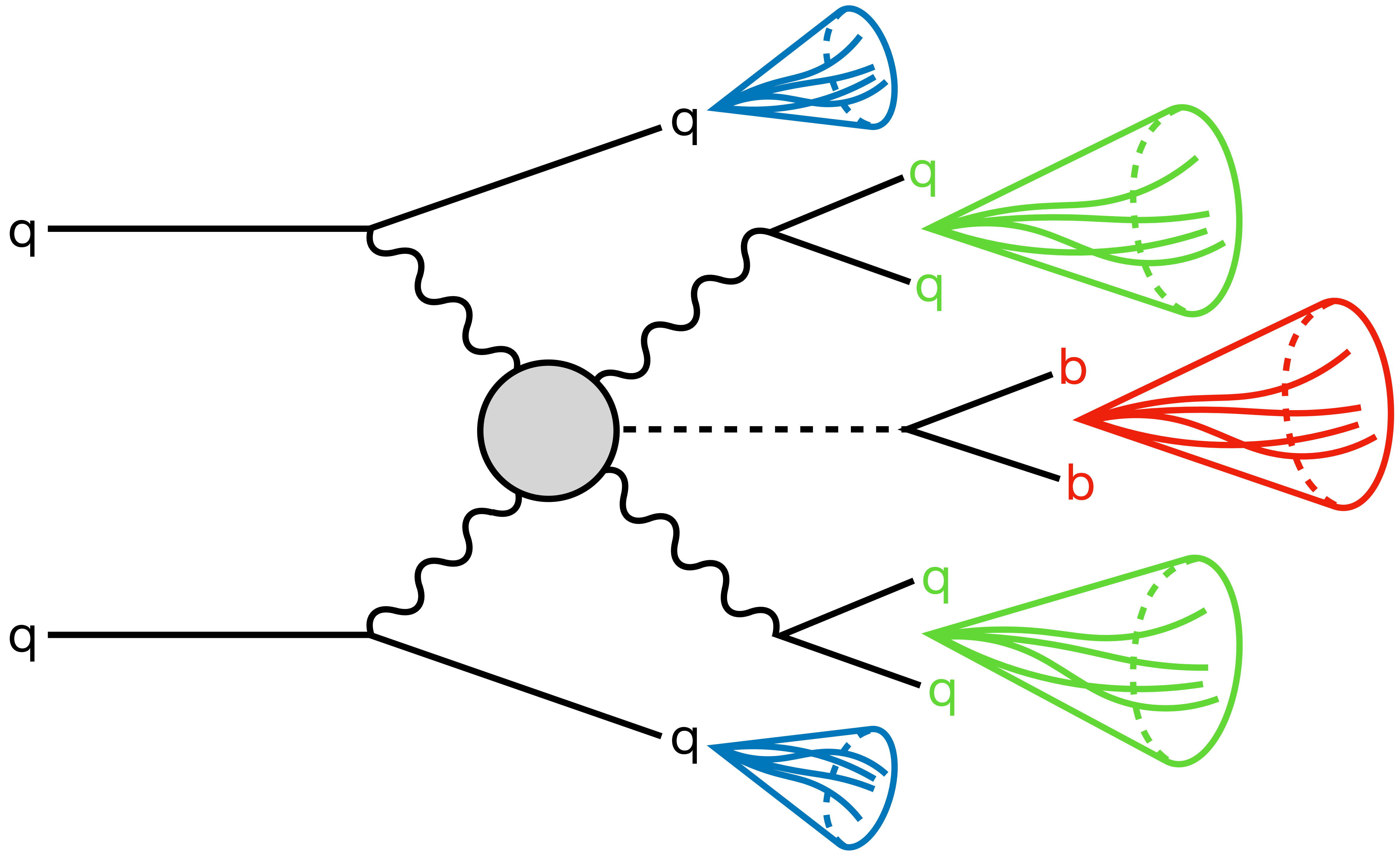
- Make loose selection ("**Preselection**") on ParticleNet scores
    - Xbb($H\rightarrow b\bar{b}$) > 0.5 and XWqq(ld $V\rightarrow qq$) > 0.3 and XWqq(tr $V\rightarrow qq$) > 0.3
- Signal region optimization is performed only on events passing the Preselection

Analysis Handles

Yields scaled to lumi×σ, rounded for readability

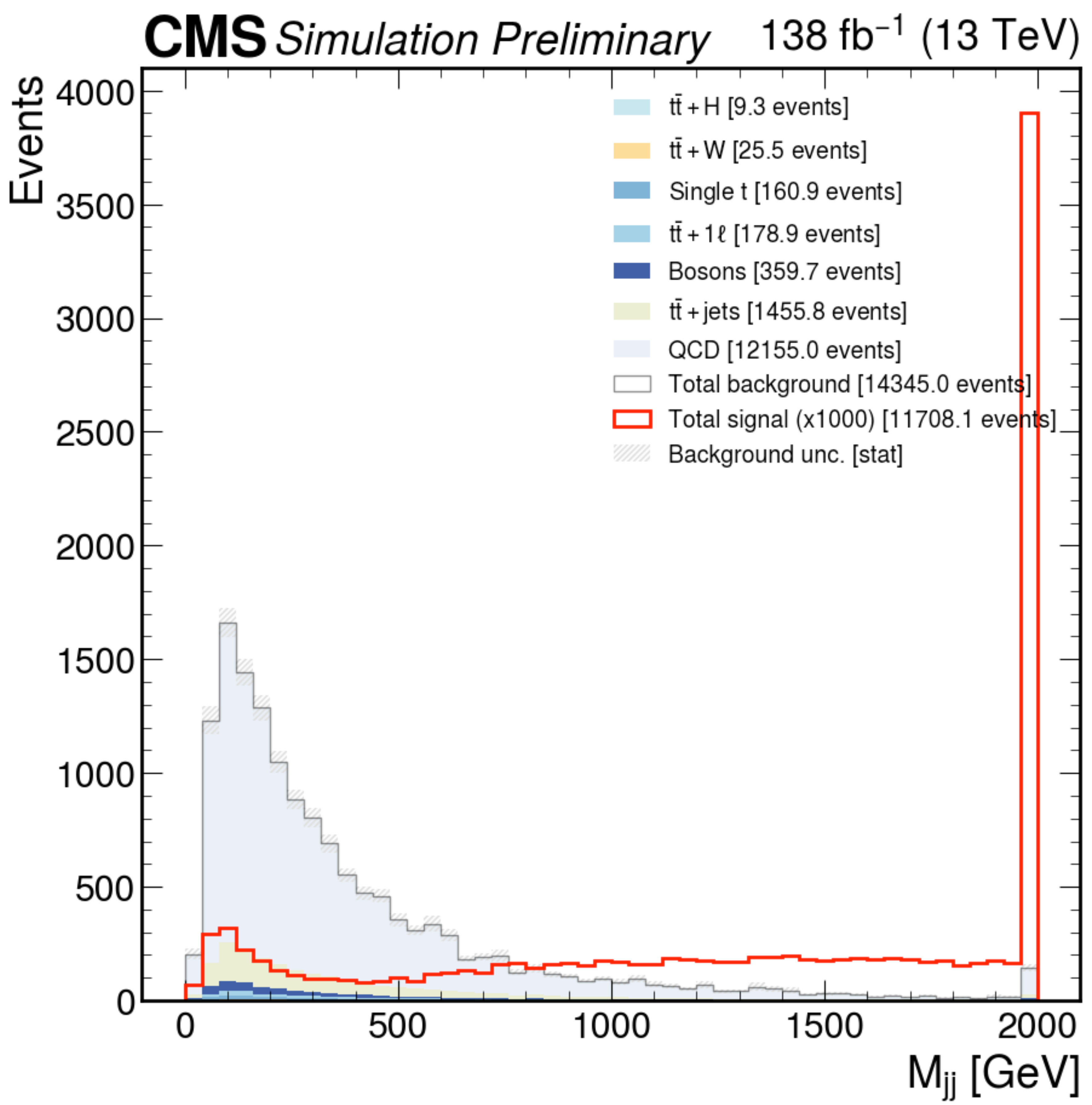
Cut	QCD	$t\bar{t}$+jets	$t\bar{t}$+1ℓ	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%
Object selection	158K	6.2K	855	59	30	478	5.1K	171K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%

- **Next:** plot the analysis handles in MC:
  - **VBS** jets M_{jj}, $|\Delta\eta_{jj}|$
  - **H→bb** fat jet M_{PNet}, p_T
  - Leading (ld) **V→qq** fat jet M_{PNet}, p_T
  - Trailing (tr) **V→qq** fat jet M_{PNet}, p_T
  - $S_T = p_T(\mathbf{H}) + p_T(\mathbf{V}) + p_T(\mathbf{V})$

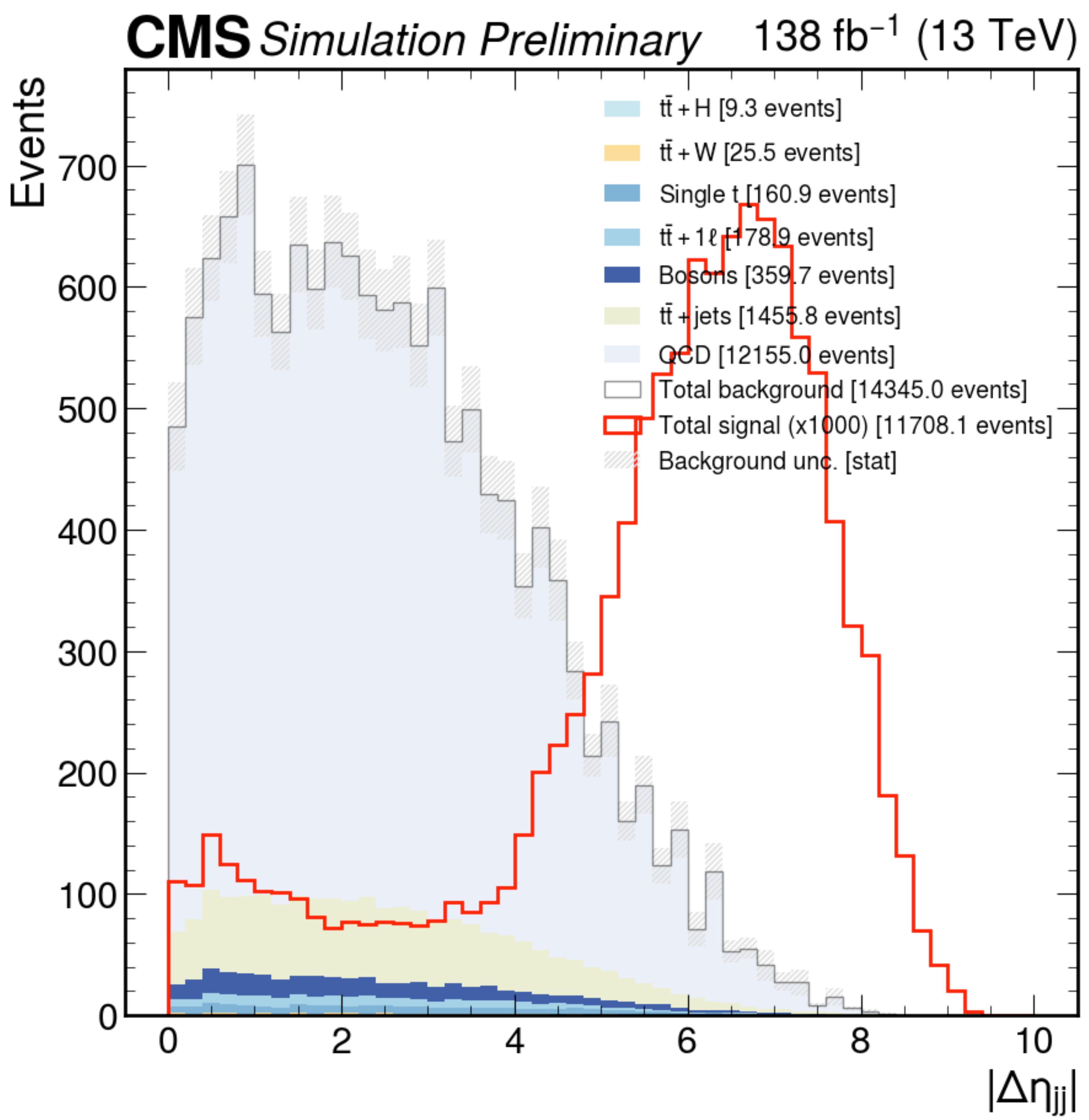


UC San Diego

Analysis Handles: VBS (Preselection)



M_{jj} = Mass(ld VBS jet p4 + tr VBS jet p4)

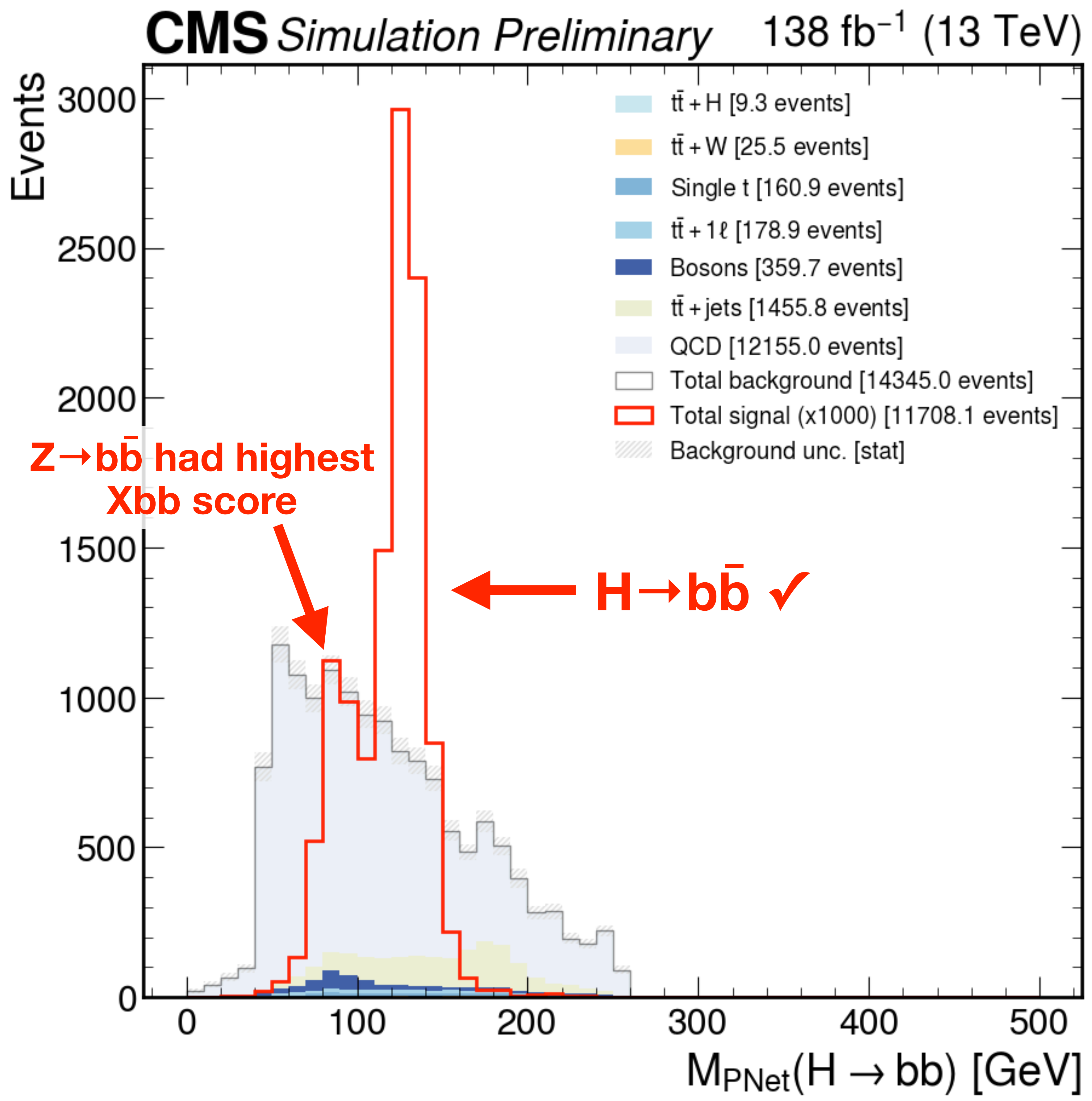


$|\Delta\eta_{jj}|$ = |ld VBS jet η - tr VBS jet η|

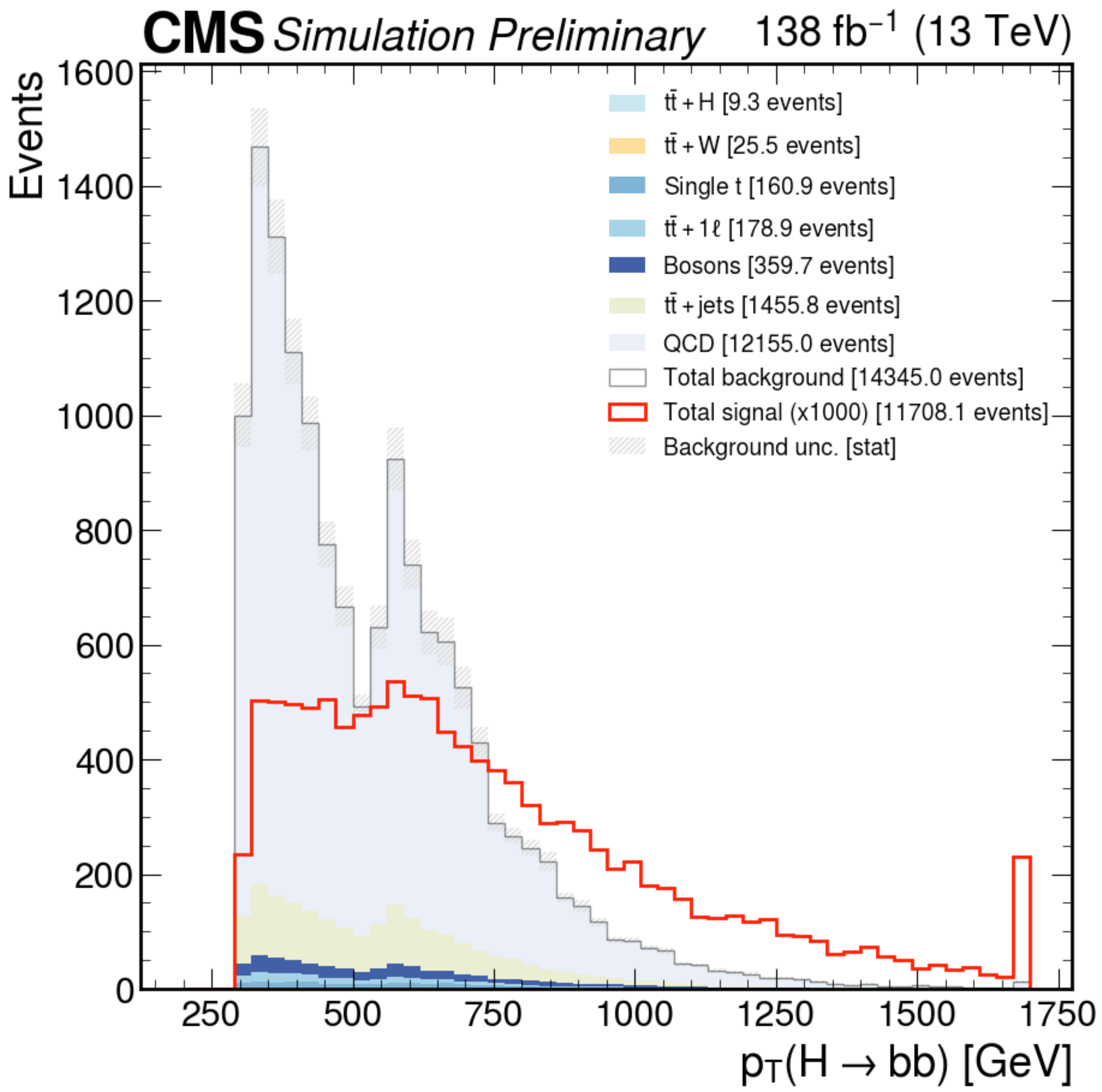
Characteristically large $\Delta\eta_{jj}$ and M_{jj} for signal (C_{2V} = 2)

UC San Diego

Analysis Handles: H→bb (Preselection)



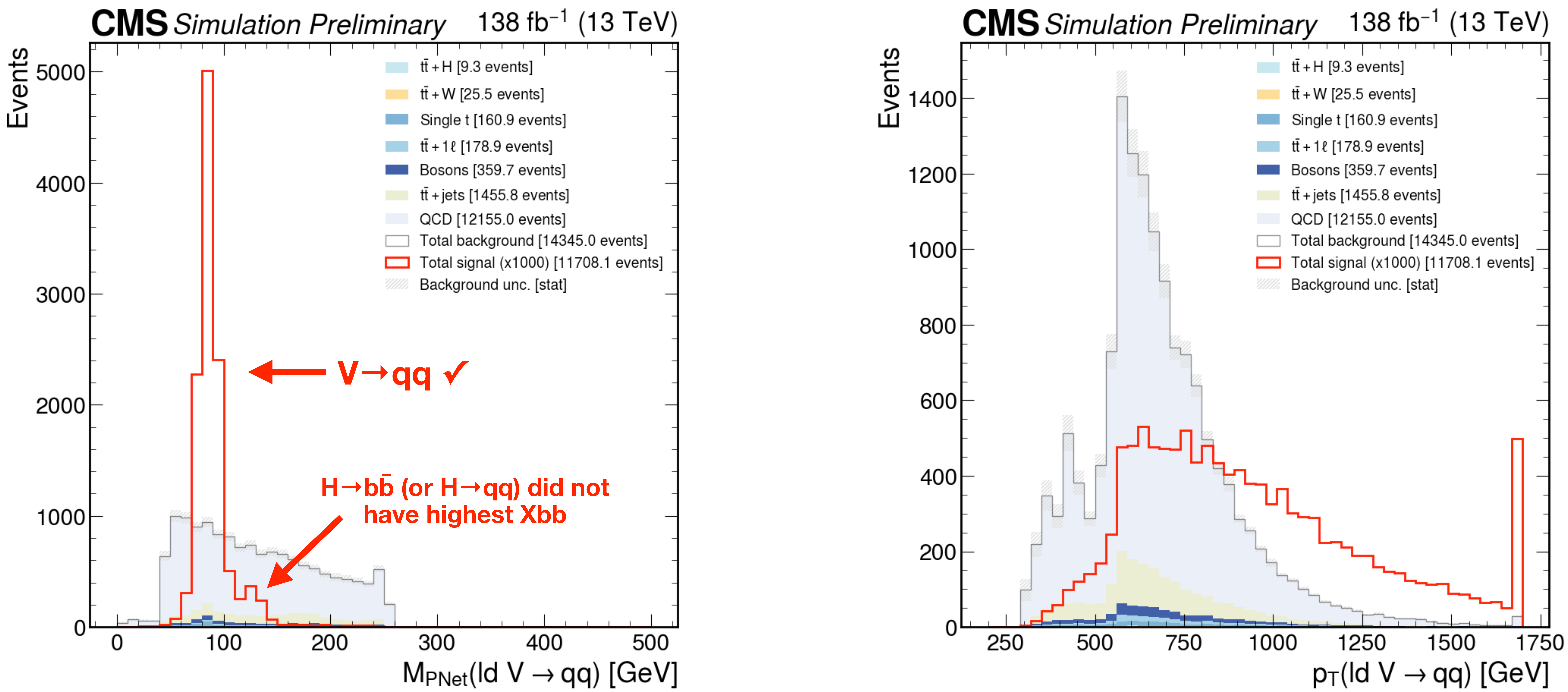
M_{PNet} = ParticleNet regressed mass



Spike at 500 GeV due to HLT threshold cut

Higgs peak in regressed mass + large p_T for signal (C2V = 2)

Analysis Handles: Id V→qq (Preselection)



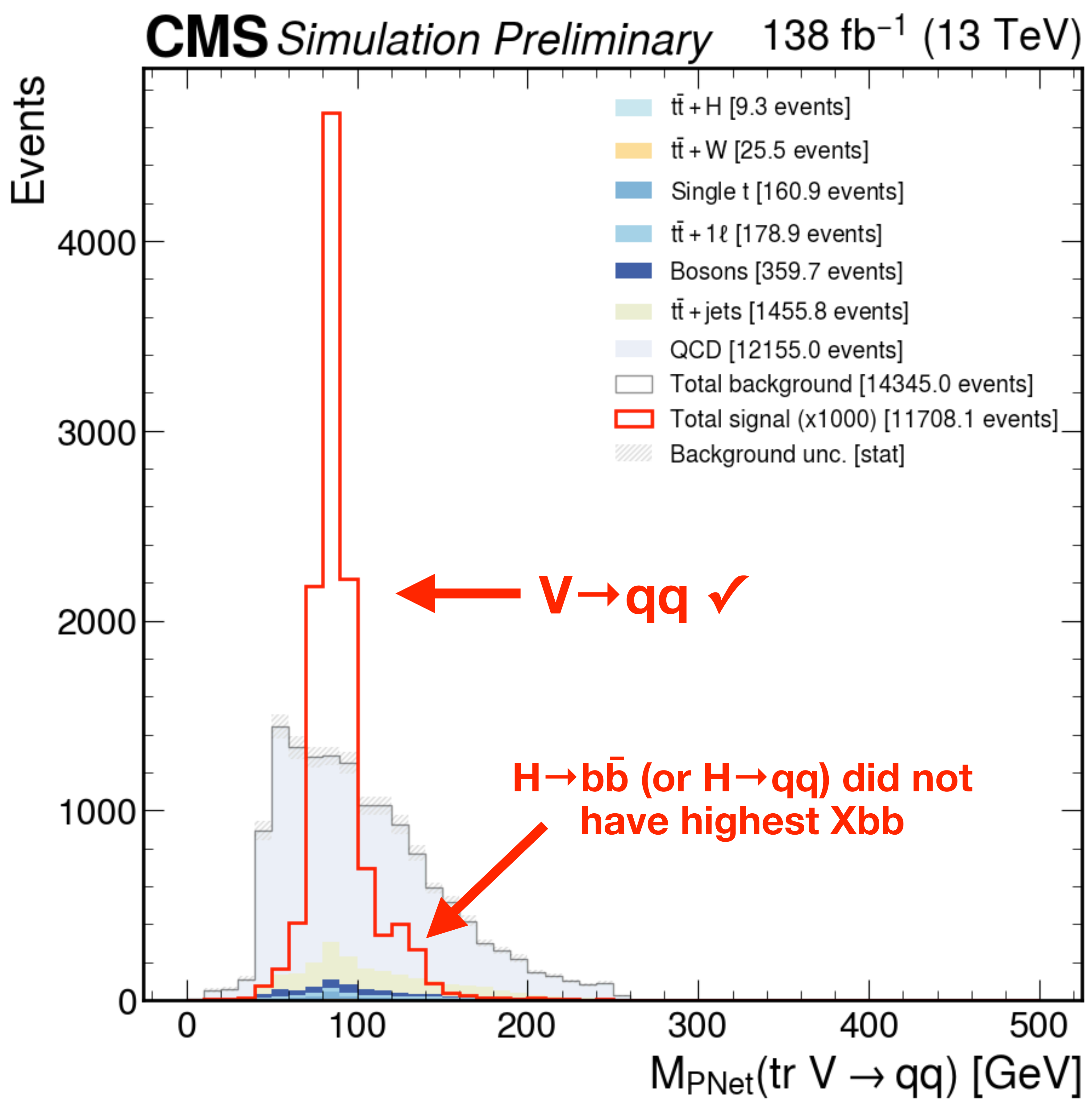
M_{PNet} = ParticleNet regressed mass

Spike at 500 GeV due to HLT threshold cut

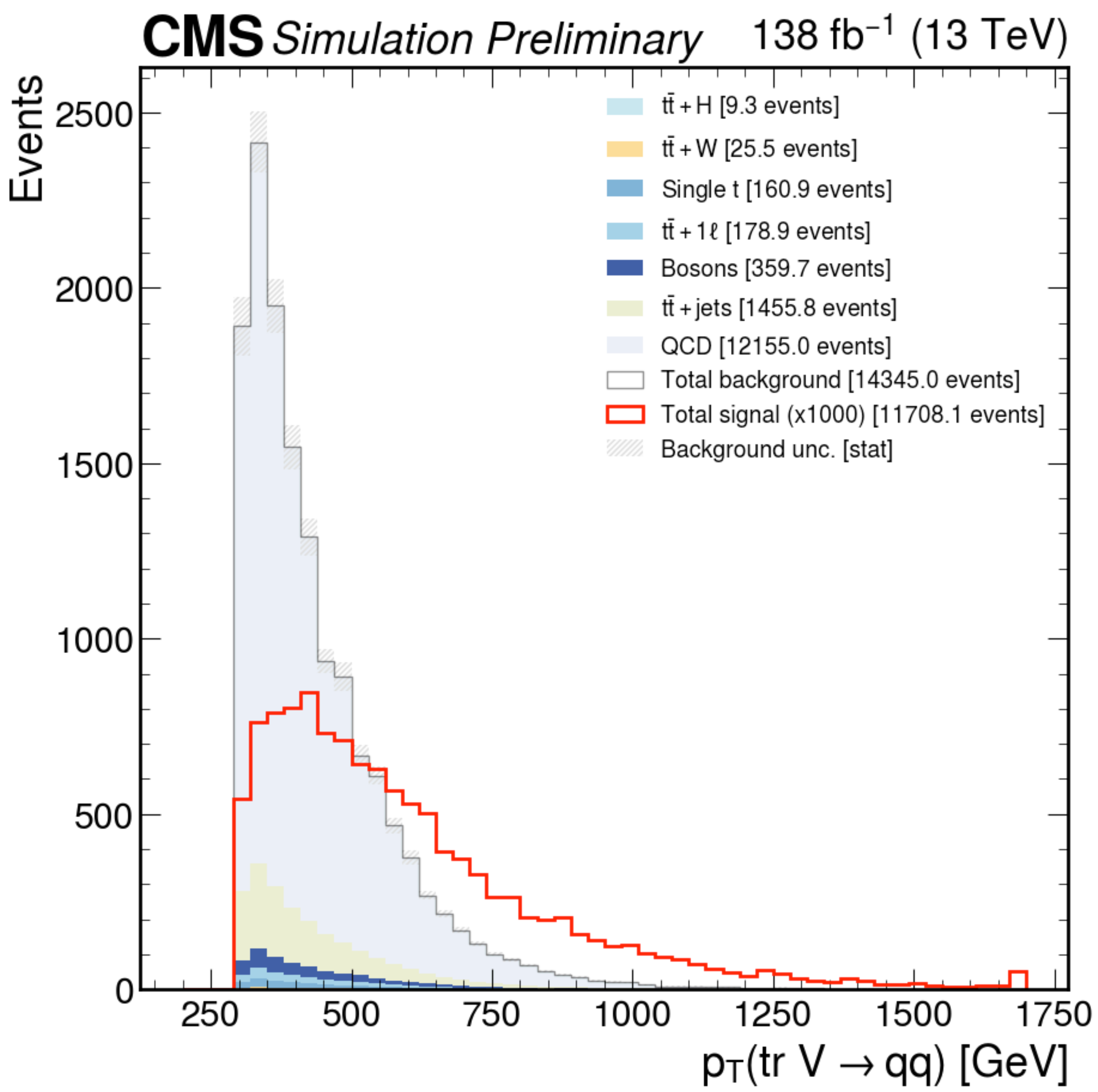
W/Z peak in regressed mass + large p_T for signal (C2V = 2)

UC San Diego

Analysis Handles: tr V→qq (Preselection)



M_{PNet} = ParticleNet regressed mass

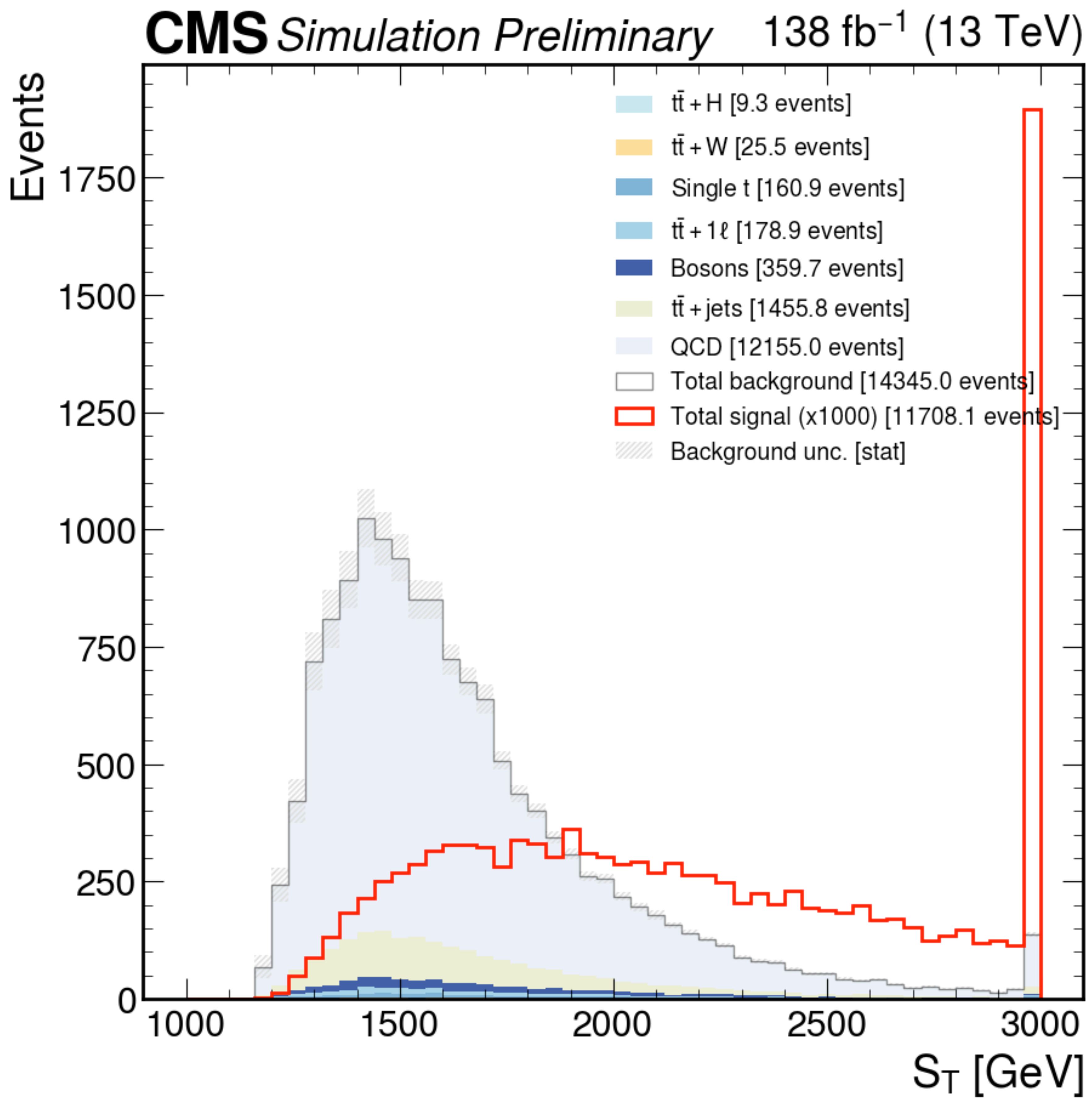


Trailing (tr) in p_T ⇒ no spike at 500 GeV from cut

W/Z peak in regressed mass + large p_T for signal (C2V = 2)

UC San Diego

Analysis Handles: S_T (Preselection)

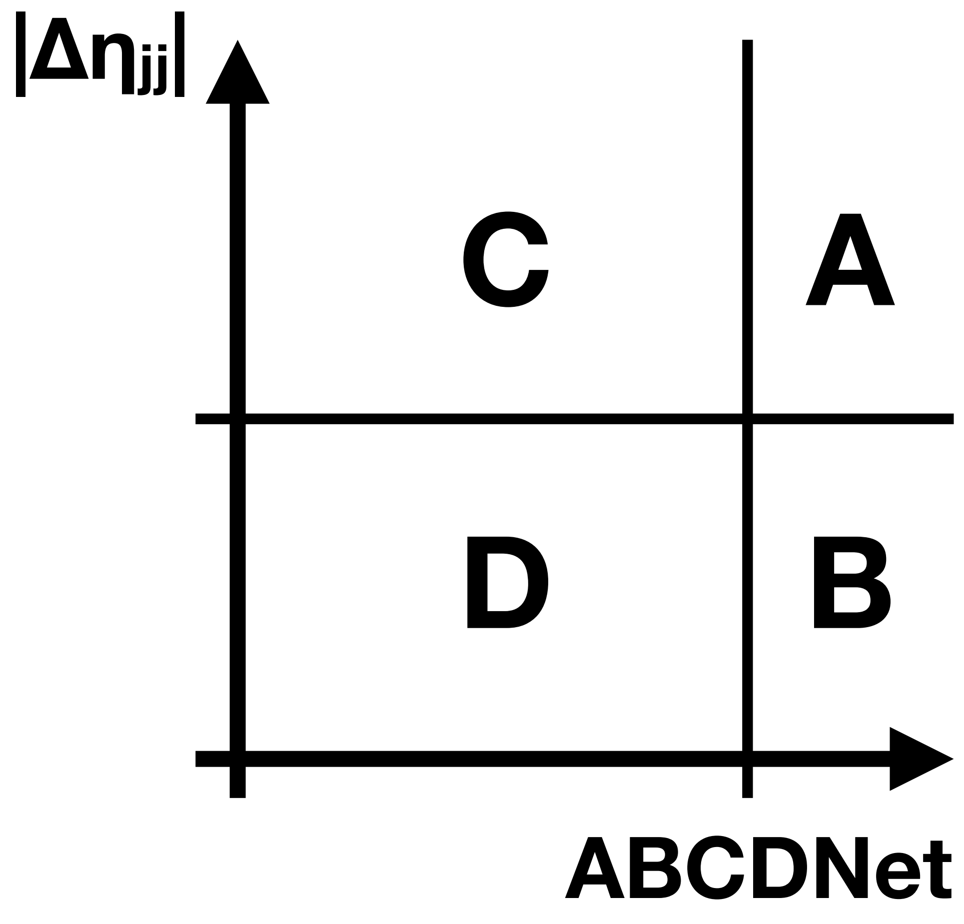


$S_T = p_T(H\rightarrow bb) + p_T(\text{ld } V\rightarrow qq) + p_T(\text{tr } V\rightarrow qq)$

Expectedly large S_T for signal (C2V = 2)

Multivariate Analysis and Background Extrapolation

Automated ABCD



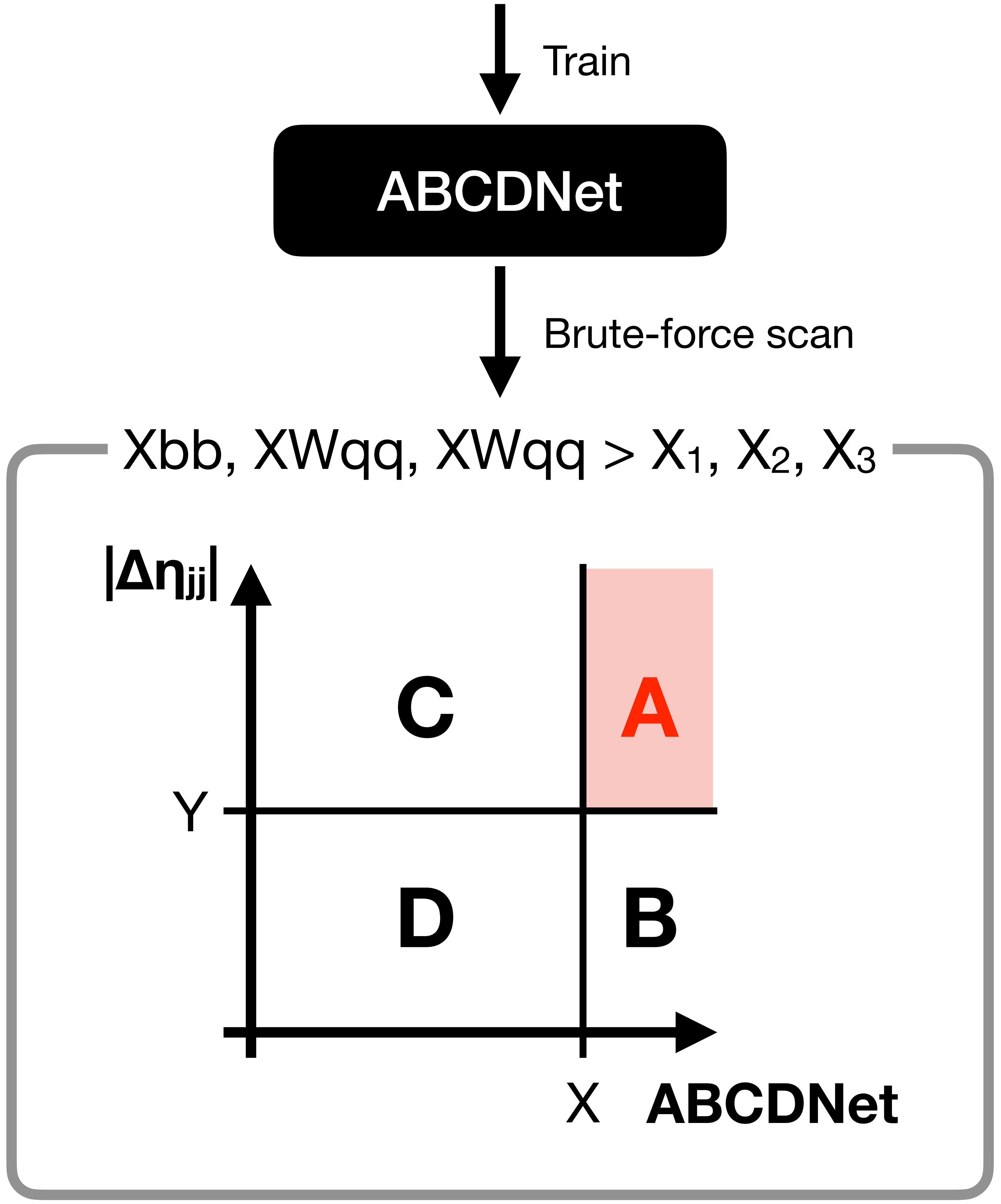
- Need the following:
    1. Performant signal region (SR)
    2. Uncorrelated “arms” to perform ABCD bkg. estimation that closes
- **Previously used BDT** as one arm, VBS cuts as the other, but **did not close (backup)**
- Enter: automated ABCD via ML (10.1103/PhysRevD.103.035021)
    - Train a **deep neural network (“ABCDNet”) as one arm,** use **$|\Delta\eta_{jj}|$ as the other**
    - **Add a decorrelation term to the loss function** that trains the network to be decorrelated from $|\Delta\eta_{jj}|$
- **Therefore, we get both (1) and (2) at the same time!**

Automated ABCD: Step-by-Step

1. Re-sample QCD ParticleNet scores for better training statistics after Preselection
2. Train ABCDNet to classify signal vs. bkg
   - Use signal and bkg events at Preselection
3. Form a **SR**: ABCDNet, ParticleNet, and $|\Delta\eta_{jj}|$ cuts
   - "Brute-force" scan of many **SR** candidates
4. Use ABCD method to predict bkg yield in **SR**
   - Test closure in MC
   - Test closure in data sidebands

Yields scaled to lumi×σ, rounded for readability

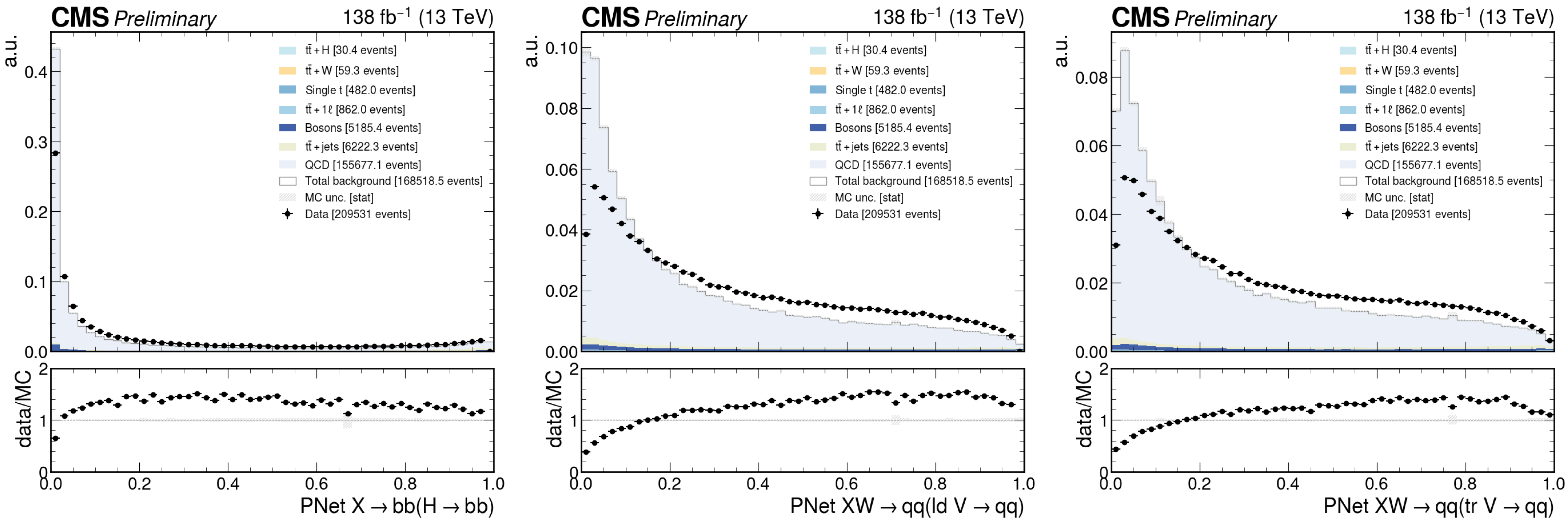
Cut	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Preselection	14K	–	12	–



UC San Diego

Step 1

Data vs. MC: ParticleNet Scores



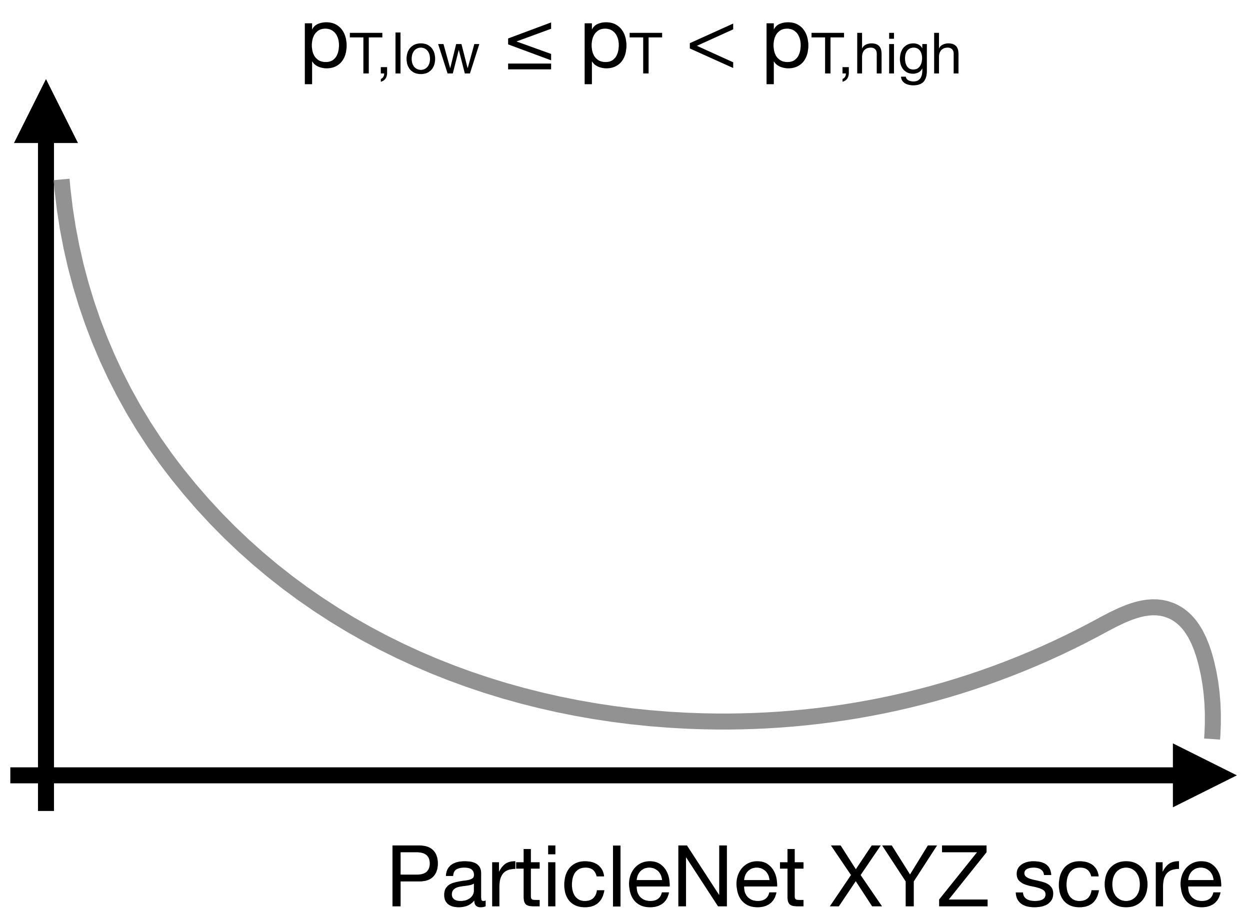
Excess MC for ParticleNet ~ 0 and roughly constant deficit MC in the tail

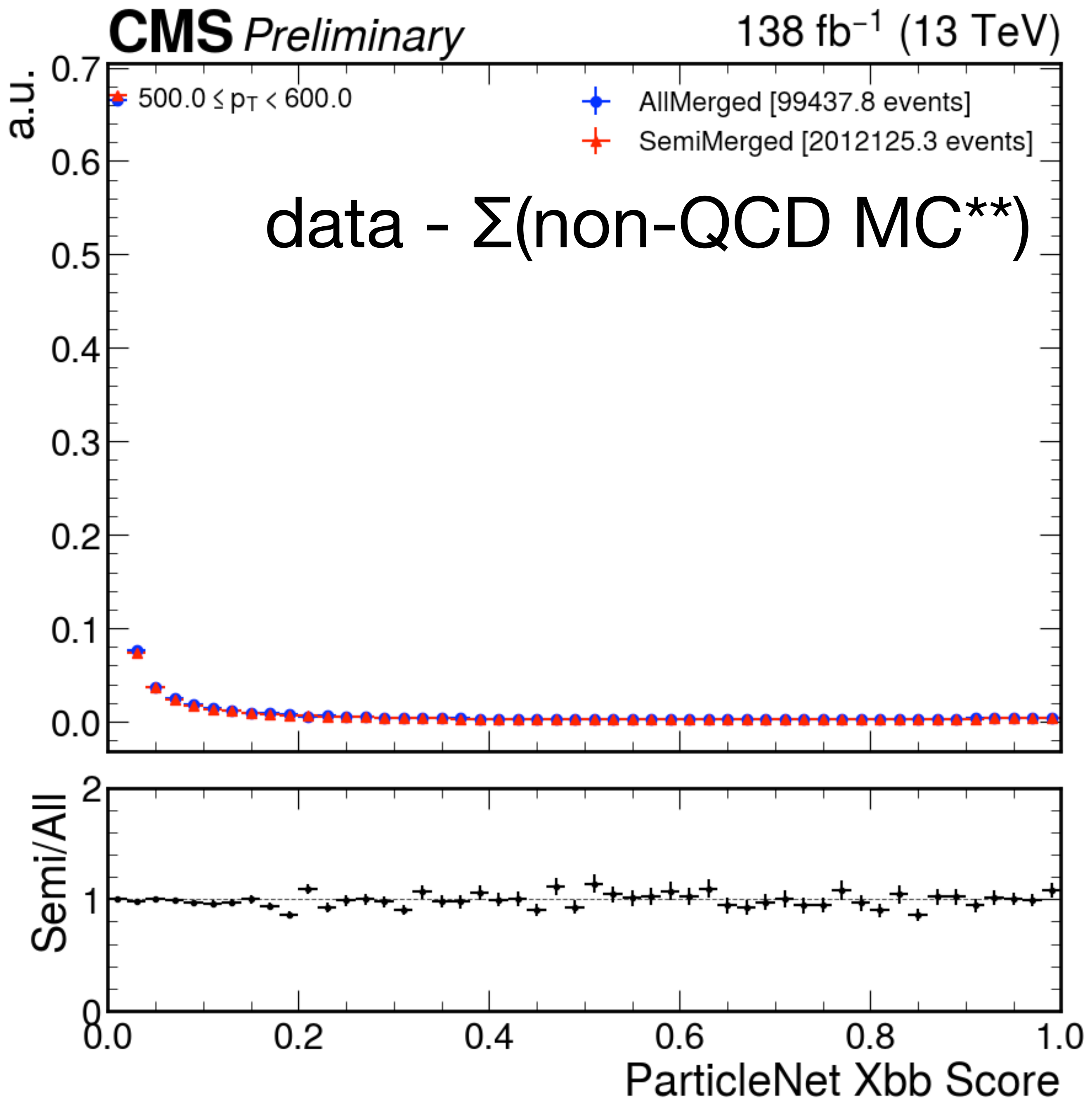
If QCD was more correctly distributed, would have more stats after Preselection

UC San Diego

QCD Corrections: Derivation

- The ParticleNet XYZ* score for a given fat jet should be fundamentally described by a probability density function
- We can approximate this PDF by
  - Plotting the ParticleNet XYZ score **in data** for every fat jet in a histogram (really several: e.g. one per p_T bin)
  - Normalizing that histogram to unity
- The PDF should be the same for fat jets in the **3 fat jet channel** (main) vs. **2 fat jet channel** ✓
- **Goal: replace MC ParticleNet scores** in 3 fat jet channel with those sampled from the 2-fat jet PDF (**from data**)





*XYZ = Xbb, XWqq, etc.
**The non-QCD MC is a small fraction of the total (backup)

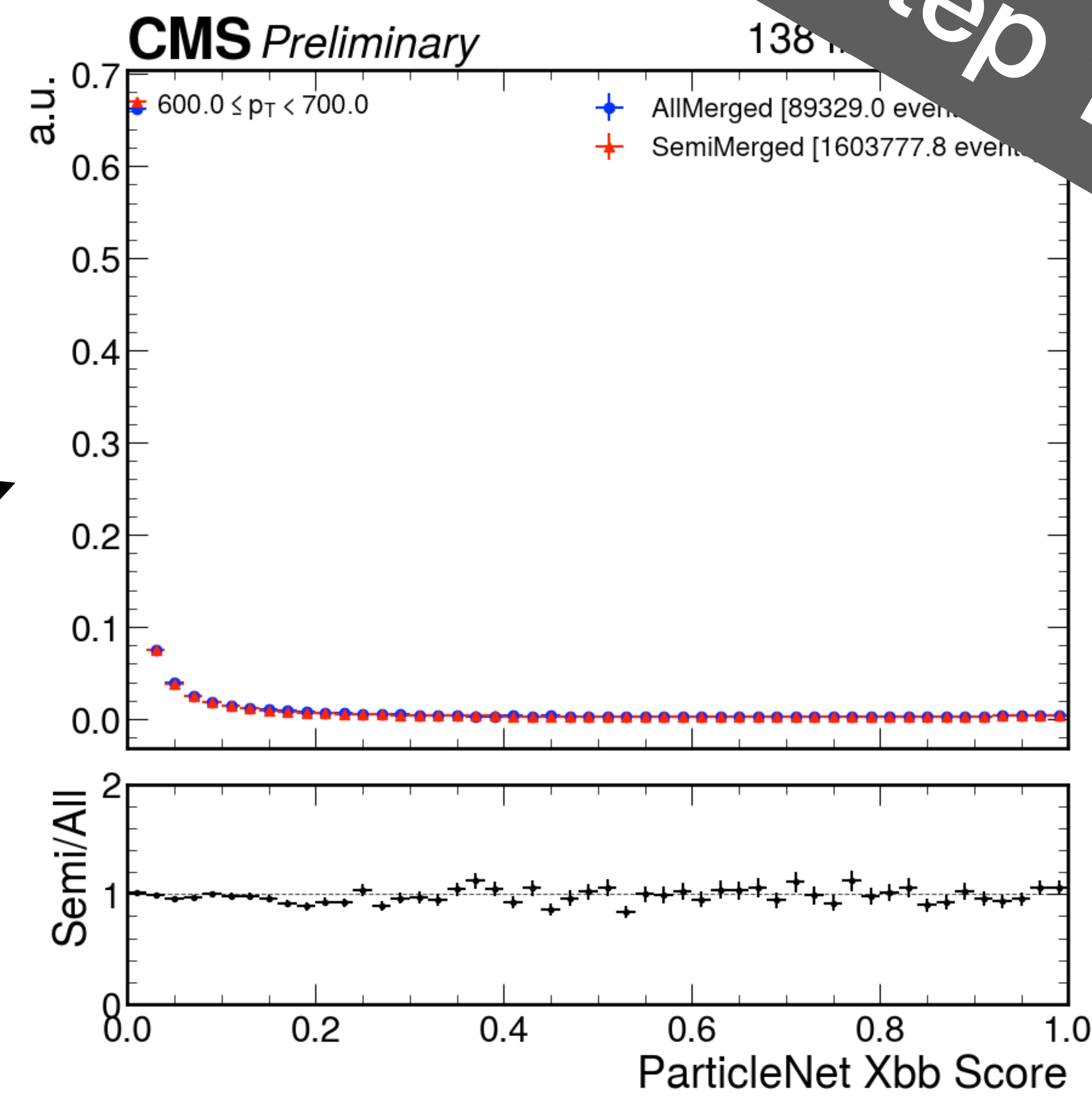
UC San Diego

Step 1

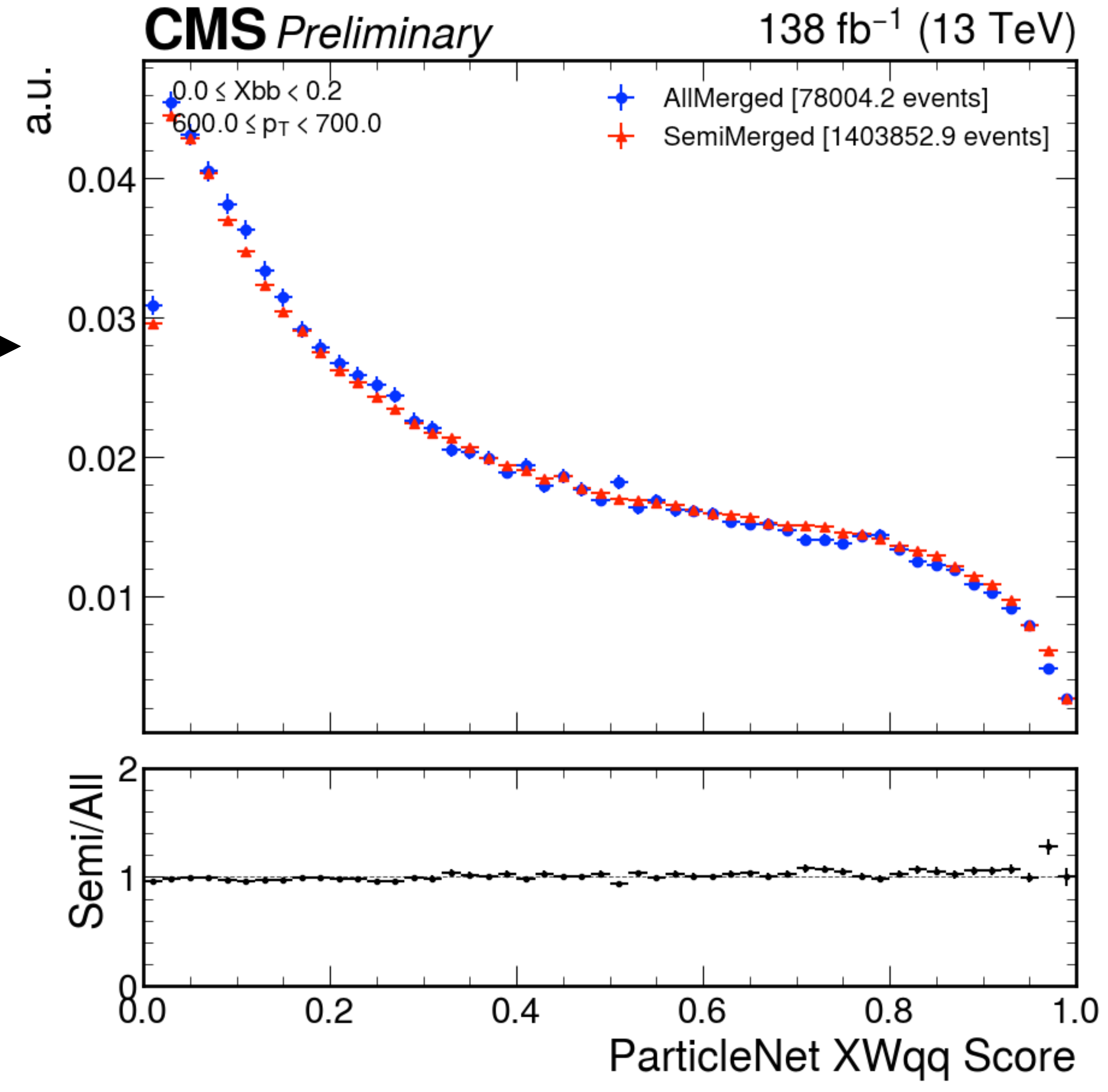
QCD Corrections: Application

- For each fat jet:
  - Get fat jet p_T (e.g. 650 GeV)
  - Get Xbb “PDF” for the appropriate p_T bin
  - **Replace** fat jet Xbb score with one randomly sampled from the **Xbb “PDF”** (e.g. 0.07)
  - Get XWqq “PDF” for the appropriate p_T, Xbb bin
  - **Replace** fat jet XWqq score with one randomly sampled from the **XWqq “PDF”** (e.g. 0.17)

Use “SemiMerged” hist



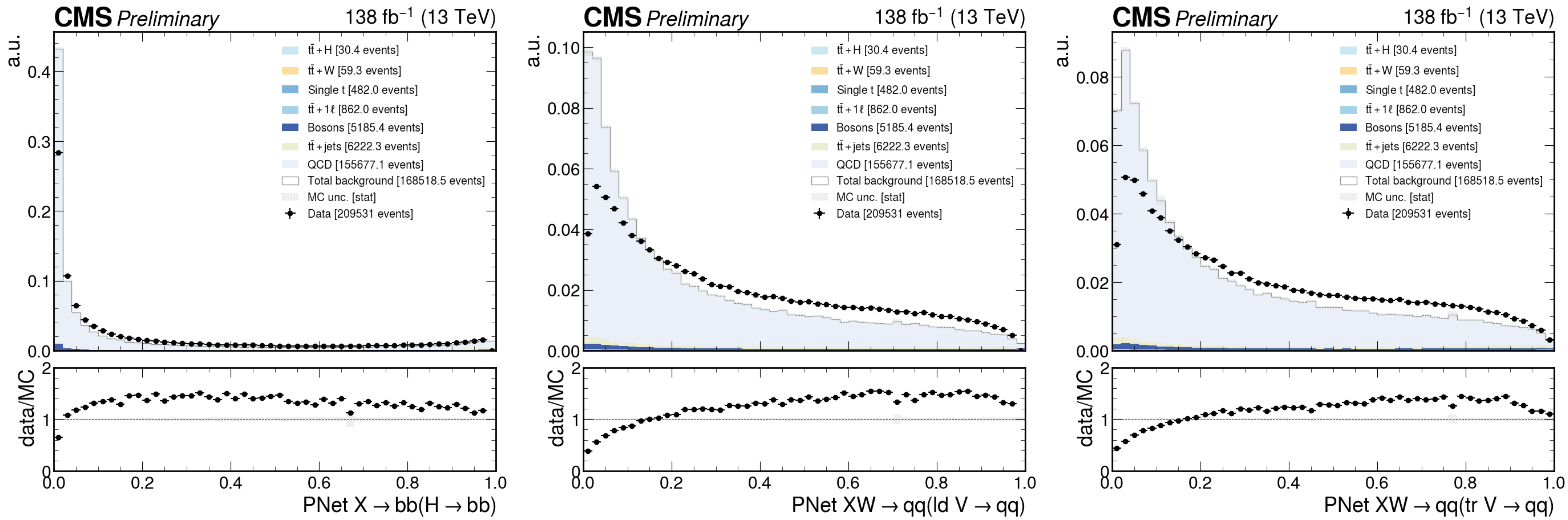
Use “SemiMerged” hist



UC San Diego

QCD Corrections: Before

Step 1



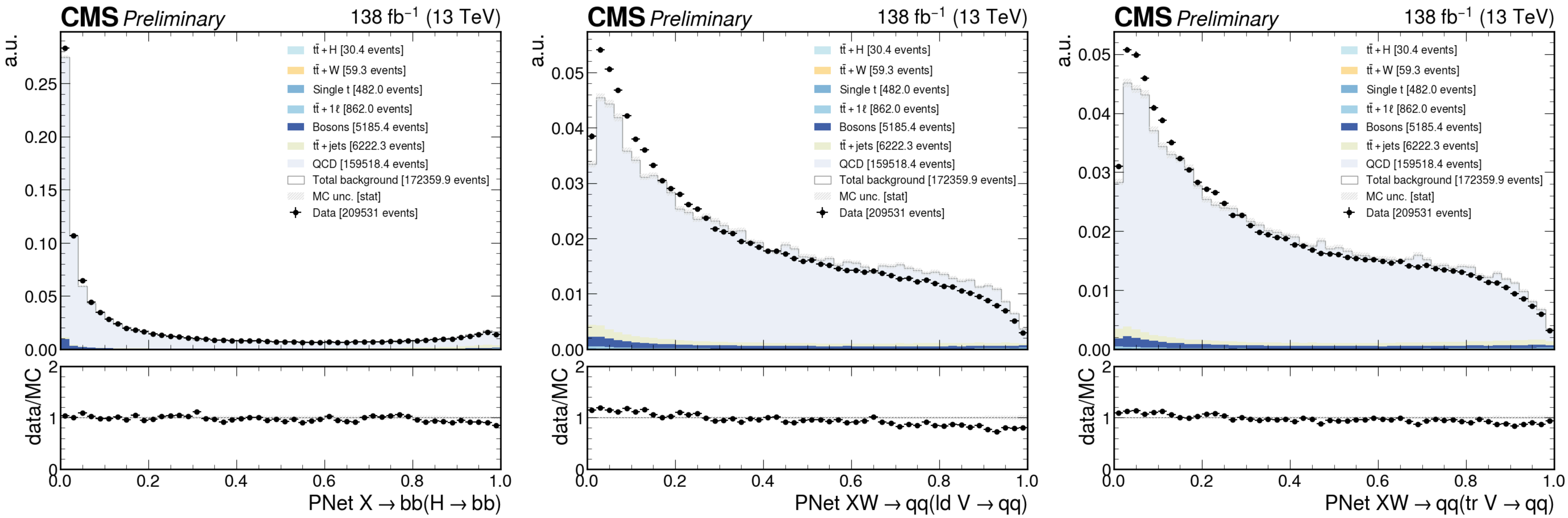
Excess MC for ParticleNet ~ 0 and roughly constant deficit MC in the tail

If QCD was more correctly distributed, would have more stats after Preselection

UC San Diego

Step 1

QCD Corrections: After



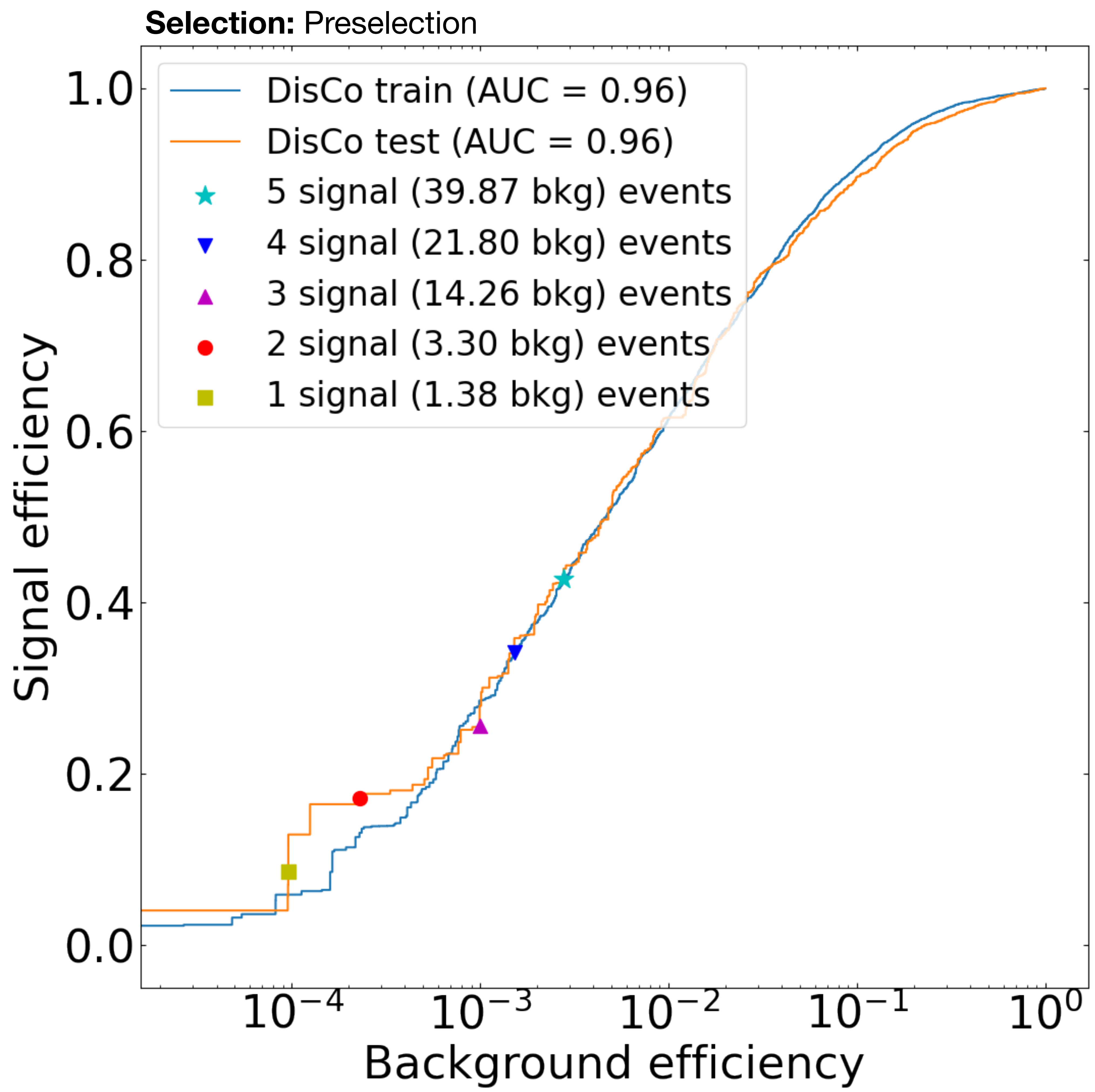
Not perfect, but agree within 20% → More QCD MC stats after Preselection!

(for optimization)

Step 2

ABCDNet: Training

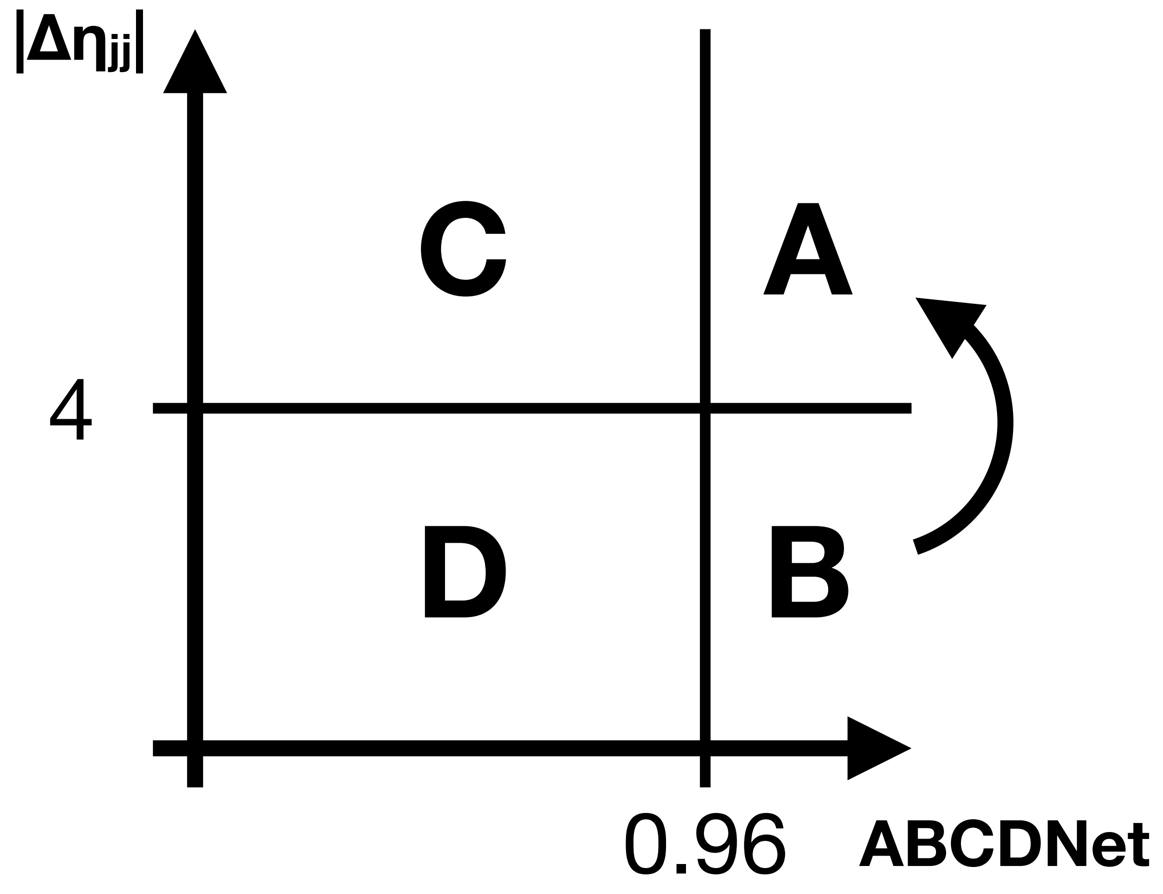
- Trained on 80% of events after Preselection
  - 20% reserved for testing
- Input features:
  - H→bb fat jet p4 (i.e. p_T, η, ϕ), M_{PNet}
  - V→qq fat jet p4 (i.e. p_T, η, ϕ), M_{PNet}
  - M_{jj}
- More training details in the backup



UC San Diego

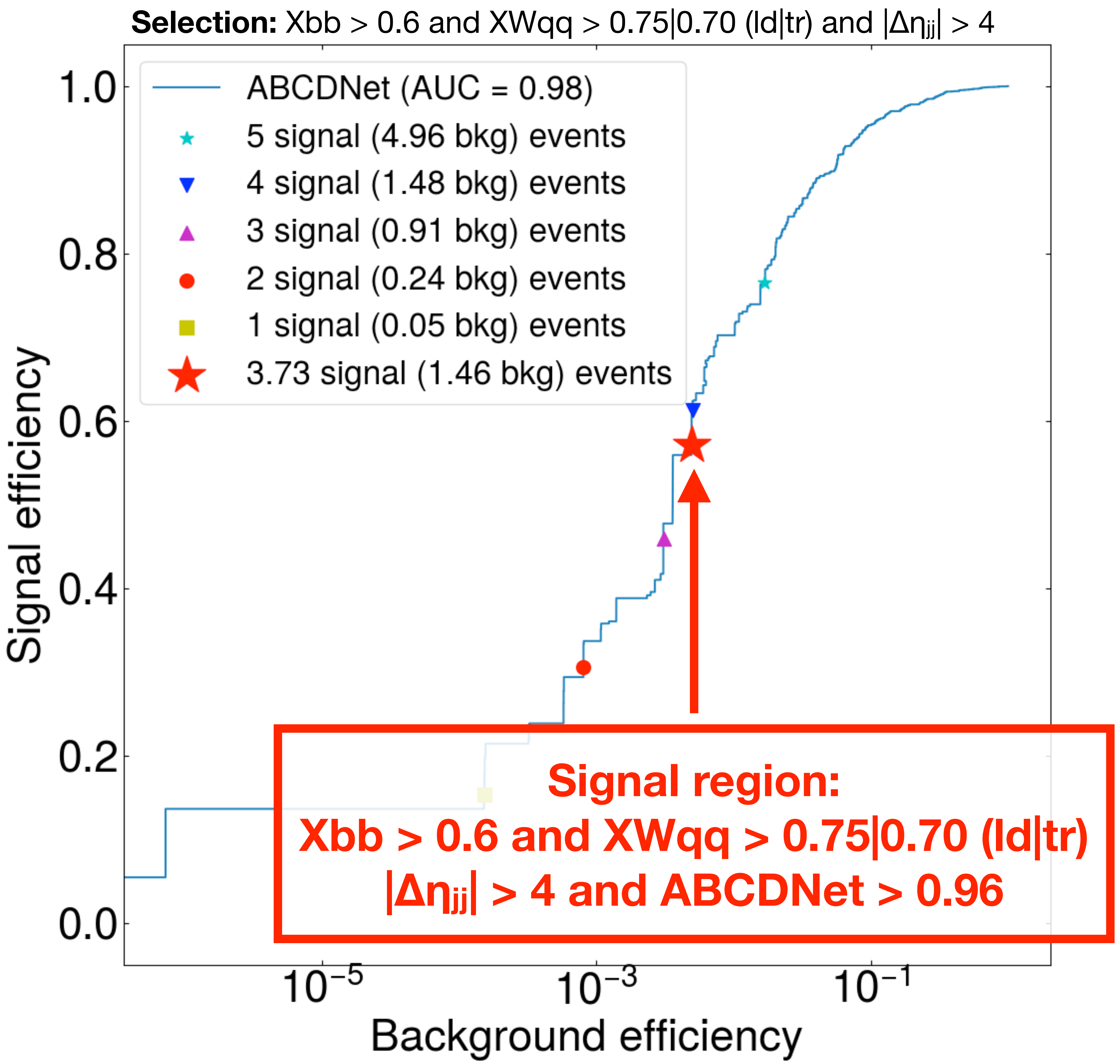
ABCDNet: Signal Region

- Perform a brute-force scan over ParticleNet taggers, ABCDNet, and $|\Delta\eta_{jj}|$
  - Tried > 7000 signal regions
  - Loosely optimize for $S/\sqrt{B}$
- Ultimately, choose **SR** (★) that works well for ABCD bkg. estimation:



Should close in MC
A = B×C/D

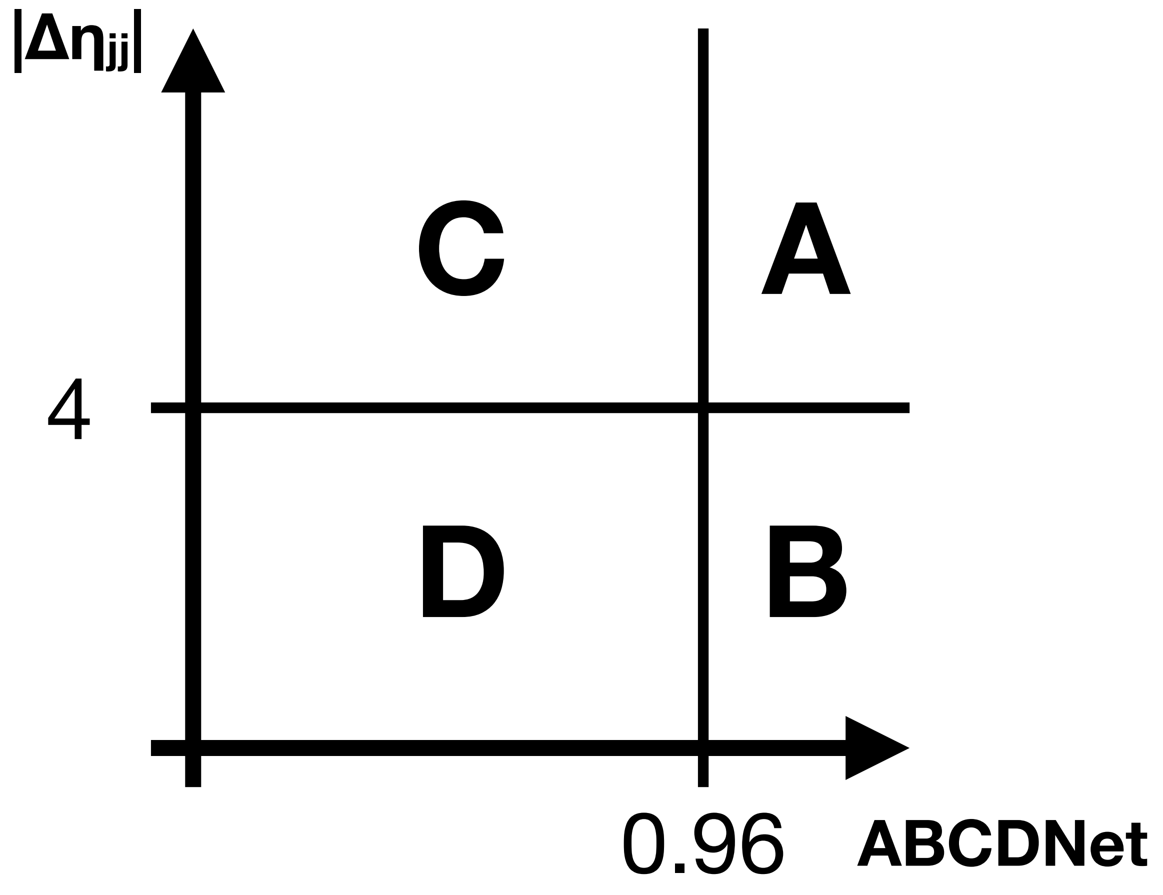
Need sufficient stats in Region B



UC San Diego

Step 4

ABCD: Configuration & MC Closure



Xbb > 0.60 and XWqq > 0.75|0.70 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
$\lvert\Delta\eta_{jj}\rvert > 4$ and ABCDNet > 0.96	A	1.46	0.54	3.73	0.05	—	—
$\lvert\Delta\eta_{jj}\rvert \leq 4$ and ABCDNet > 0.96	B	5.12	1.06	0.48	0.02	4	2.00
$\lvert\Delta\eta_{jj}\rvert > 4$ and ABCDNet ≤ 0.96	C	292.88	25.89	2.80	0.05	280	16.73
$\lvert\Delta\eta_{jj}\rvert \leq 4$ and ABCDNet ≤ 0.96	D	1012.41	41.07	0.63	0.02	1201	34.66

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = 1.48 \pm 0.34$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = 0.93 \pm 0.47$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{1.47}\sqrt{\left(\frac{0.54}{1.46}\right)^2 + \left(\frac{0.34}{1.48}\right)^2} = 29.5\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 50.4\%$$

Final Result

Expected sig. 3.73 ± 0.05

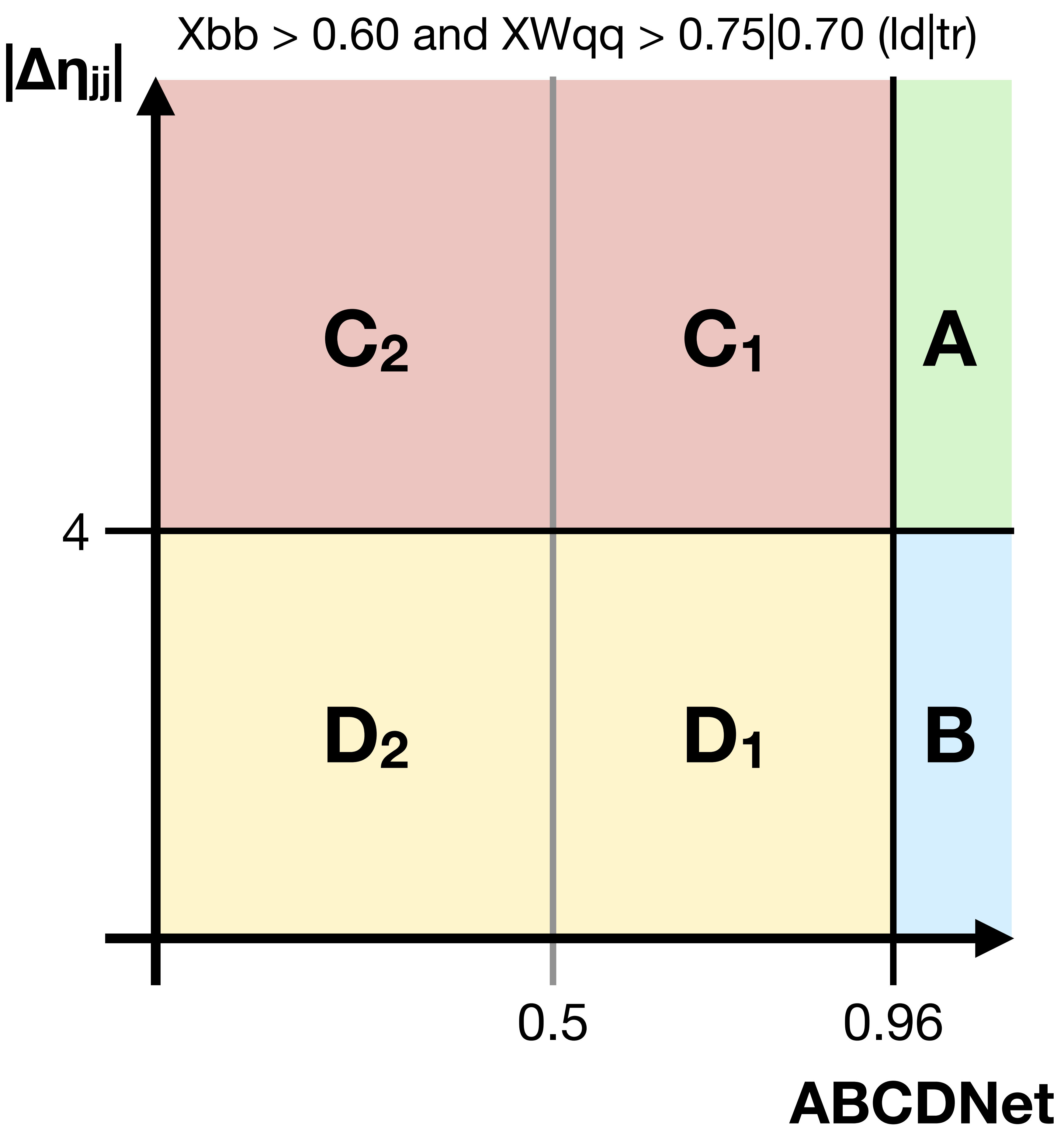
Predicted bkg. 0.93 ± 0.47 ± 0.28

stat. *syst.*

Predicted significance (S/√B) is passable and the method closes in MC

ABCD: Closure in Data

Step 4



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**1.46**	0.54	**3.73**	0.05	–	–
B	5.12	1.06	0.48	0.02	4	2.00
C	292.88	25.89	2.80	0.05	280	16.73
D	1012.4	41.07	0.63	0.02	1201	34.66

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**1.46**	0.54	**3.73**	0.05	–	–
B	5.12	1.06	0.48	0.02	4	2.00
C_1	17.17	2.46	2.29	0.04	–	–
D_1	81.2	9.25	0.47	0.02	70	8.37

C1 is signal polluted
⇒ can't unblind

$$A^{pred} = B \times \frac{C}{D} = \left[\begin{array}{ll} \mathbf{1.48 \pm 0.34} & \text{(MC)} \\ 0.93 \pm 0.47 & \text{(Data)} \end{array} \right.$$

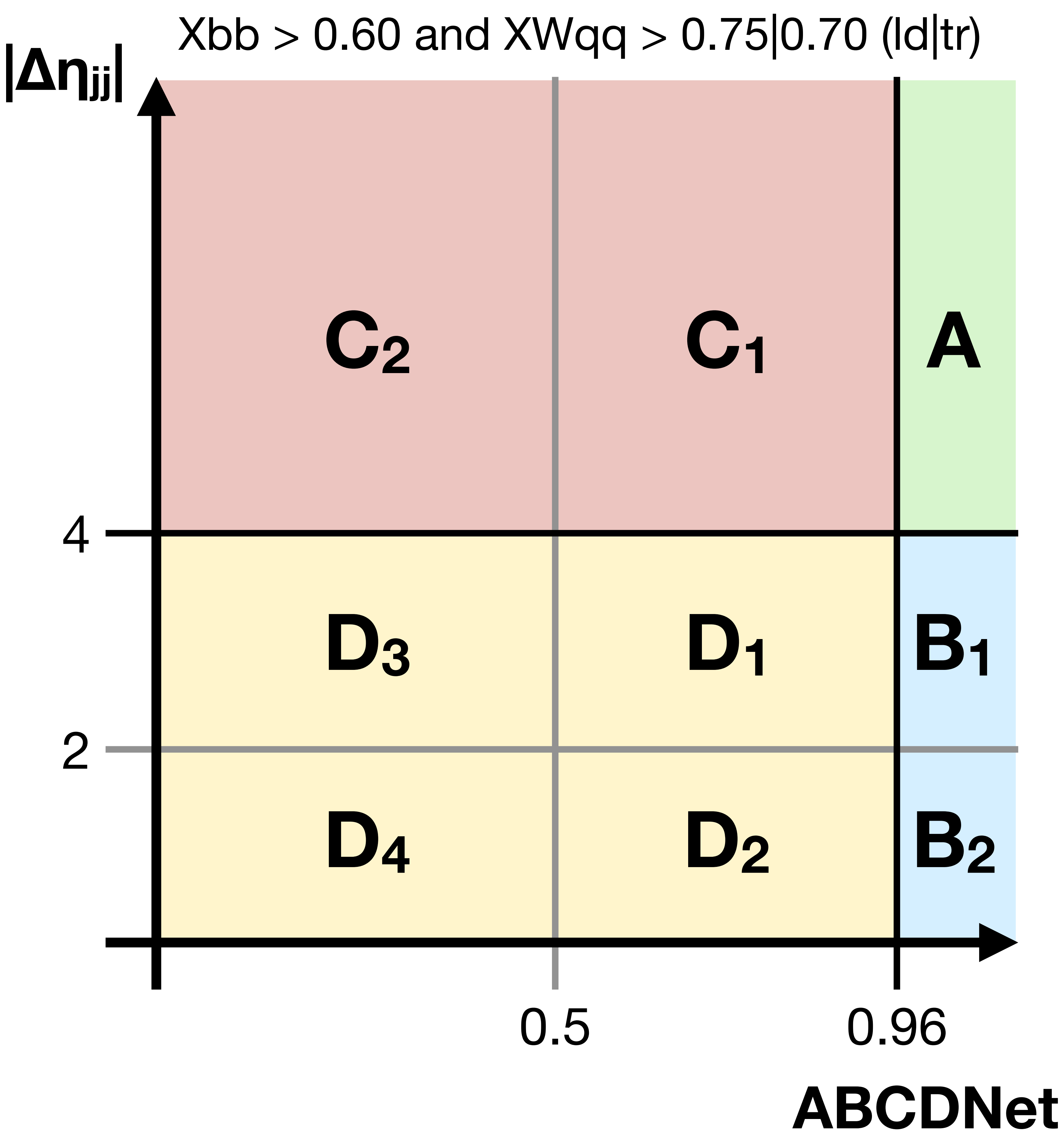
✓

$$A^{pred} = B \times \frac{C_1}{D_1} = \left[\begin{array}{ll} \mathbf{1.08 \pm 0.30} & \text{(MC)} \\ ? \pm ? & \text{(Data)} \end{array} \right.$$

ABCD works well with data in sidebands ⇒ method is valid!

Step 4

ABCD: Closure in Data



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**1.46**	0.54	**3.73**	0.05	–	–
B	5.12	1.06	0.48	0.02	4	2.00
C	292.88	25.89	2.80	0.05	280	16.73
D	1012.4	41.07	0.63	0.02	1201	34.66

$$A^{pred} = B \times \frac{C}{D} = \left[\begin{array}{l} 1.48 \pm 0.34 \text{ (MC)} \\ \mathbf{0.93 \pm 0.47} \text{ (Data)} \end{array}\right.$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	2.90	0.81	0.20	0.01	2	1.41
B_2	2.22	0.68	0.28	0.01	2	1.41
D_1	33.28	3.80	0.20	0.01	37	6.08
D_2	47.87	8.44	0.27	0.01	33	5.74

$$B_1^{pred} = B_2 \times \frac{D_1}{D_2} = \mathbf{2.24 \pm 1.67} \text{ (Data) ✓}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	33.28	3.80	0.20	0.01	37	6.08
D_2	47.87	8.44	0.27	0.01	33	5.74
D_3	441.29	28.47	0.07	0.01	540	23.24
D_4	489.96	28.12	0.10	0.01	591	24.31

$$D_1^{pred} = D_2 \times \frac{D_3}{D_4} = \mathbf{30.15 \pm 5.55} \text{ (Data) ✓}$$

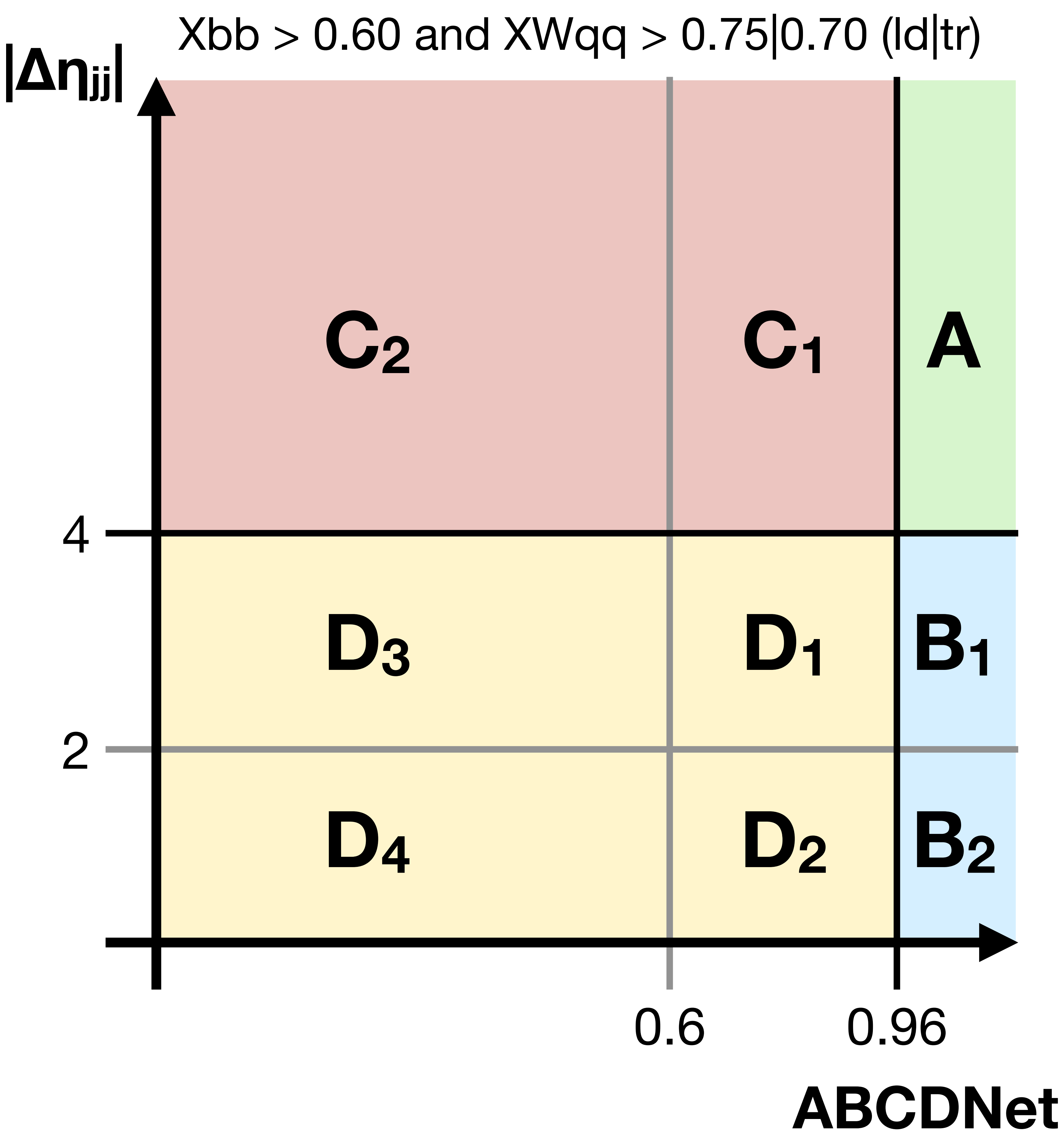
Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	17.17	2.46	2.29	0.04	–	–
D_1	33.28	3.80	0.20	0.01	37	6.08
C_2	275.71	25.78	0.51	0.02	264	16.25
D_3	441.29	28.47	0.07	0.01	540	23.24

$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} =$$ **C1 is signal polluted ⇒ can't unblind**

ABCD works well with data in sidebands ⇒ method is valid!

Step 4

ABCD: Closure in Data



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**1.46**	0.54	**3.73**	0.05	–	–
B	5.12	1.06	0.48	0.02	4	2.00
C	292.88	25.89	2.80	0.05	280	16.73
D	1012.4	41.07	0.63	0.02	1201	34.66

$$A^{\text{pred}} = B \times \frac{C}{D} = \begin{cases} 1.48 \pm 0.34 \ \text{(MC)} \\ \mathbf{0.93 \pm 0.47} \ \text{(Data)} \end{cases}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	2.90	0.81	0.20	0.01	**2**	1.41
B_2	2.22	0.68	0.28	0.01	2	1.41
D_1	27.26	3.48	0.18	0.01	30	5.48
D_2	37.87	8.14	0.24	0.01	26	5.10

$$B_1^{\text{pred}} = B_2 \times \frac{D_1}{D_2} = \mathbf{2.31 \pm 1.75} \ \text{(Data)} \ \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	27.26	3.48	0.18	0.01	**30**	5.48
D_2	37.87	8.14	0.24	0.01	26	5.10
D_3	447.31	28.51	0.09	0.01	547	23.39
D_4	499.96	28.21	0.12	0.01	598	24.45

$$D_1^{\text{pred}} = D_2 \times \frac{D_3}{D_4} = \mathbf{23.78 \pm 4.87} \ \text{(Data)} \ \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	14.91	2.33	2.15	0.04	–	–
D_1	27.26	3.48	0.18	0.01	30	5.48
C_2	277.97	25.79	0.64	0.02	269	16.40
D_3	447.31	28.51	0.09	0.01	547	23.39

$$C_1^{\text{pred}} = D_1 \times \frac{C_2}{D_3} =$$ **C1 is signal polluted ⇒ can't unblind**

ABCD works well with data in sidebands ⇒ method is valid!

UC San Diego

Summary

- We apply a correction to the QCD MC so it better resembles data for optimization
- We train a neural network (“ABCDNet”) to classify signal vs. bkg and be decorrrelated from a target variable ($\Delta\eta_{jj}$)
- We use ABCDNet and $\Delta\eta_{jj}$ to do a fully data-driven ABCD background estimation
  - We see good closure in MC and in data in various sidebands
- We obtain the following expected result:

Expected sig.	**3.73 ± 0.05**		
Predicted bkg.	**0.93**	**± 0.47**	**± 0.28**
		stat.	*syst.*

UC San Diego

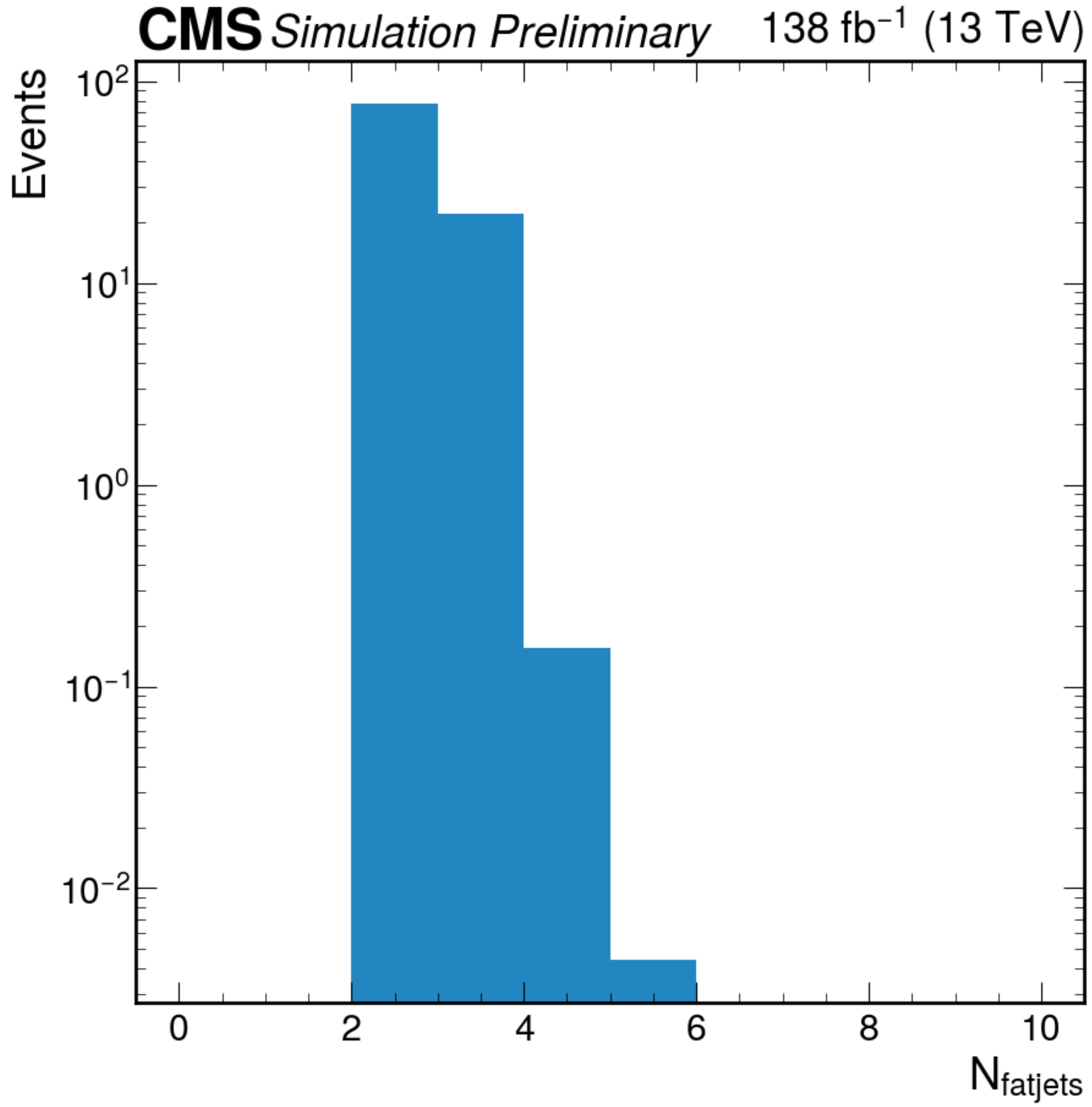
Next Steps

- Follow up on any questions about the analysis strategy presented here
- Derive a full set of systematic uncertainties on the expected signal yield:
  - Scale factors: Lepton ID, ParticleNet (Xbb, XWqq), trigger efficiency, Pileup ID
  - JEC/JER, MET
  - Pileup reweighting
  - PDF
  - Renormalization and factorization scales
  - Luminosity
- Produce an exclusion interval for C_{2V}

Backup: Analysis Overview

$N_{fatjets}$

- Plotting sum of all VBS VVH signals here
- Selections applied: skim, HLT triggers, gen-level H, W, Z decay hadronically
- Practically 0 events with more than three fatjets

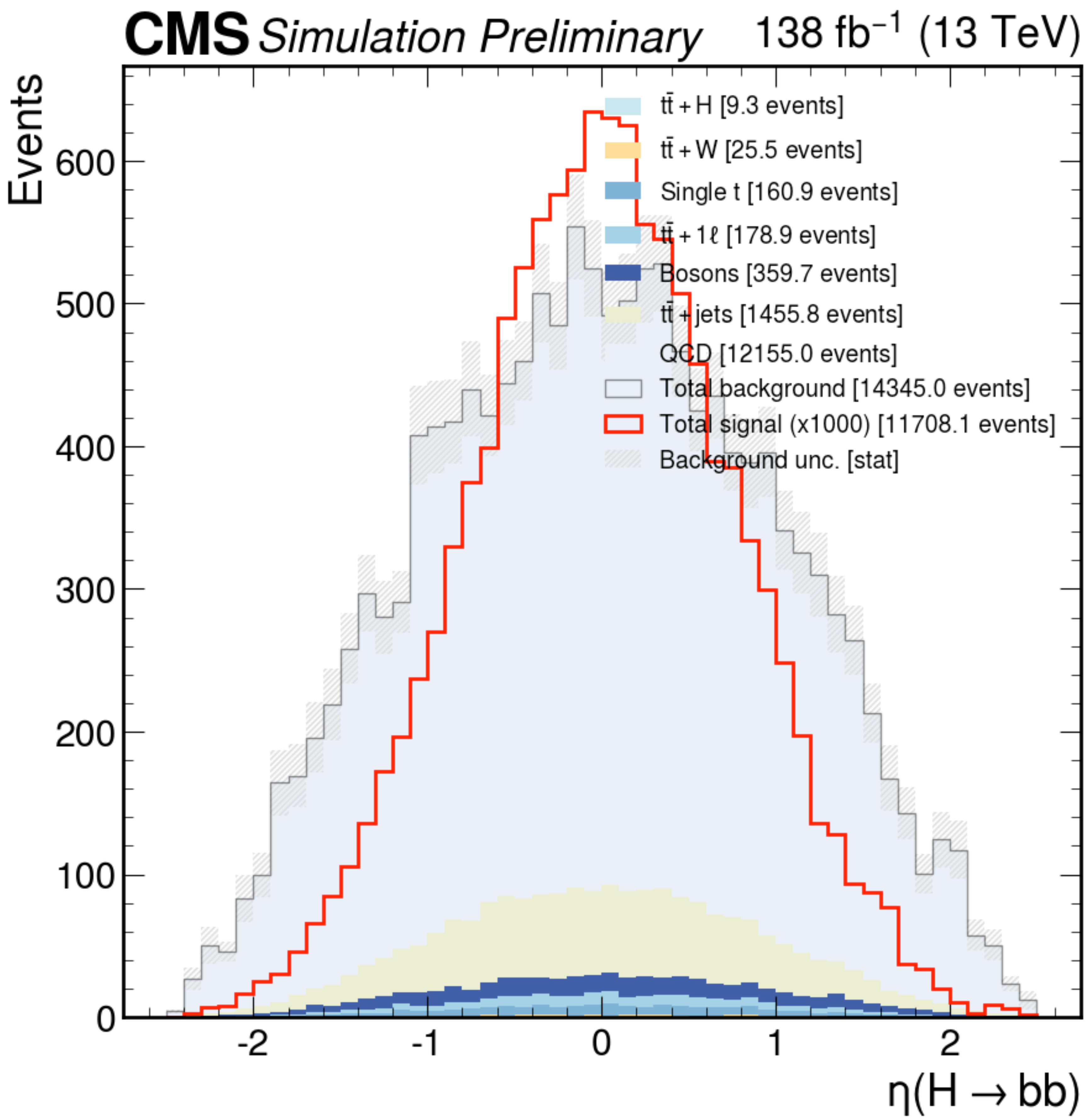


Full Cutflow

Yields scaled to lumi×σ, rounded for readability

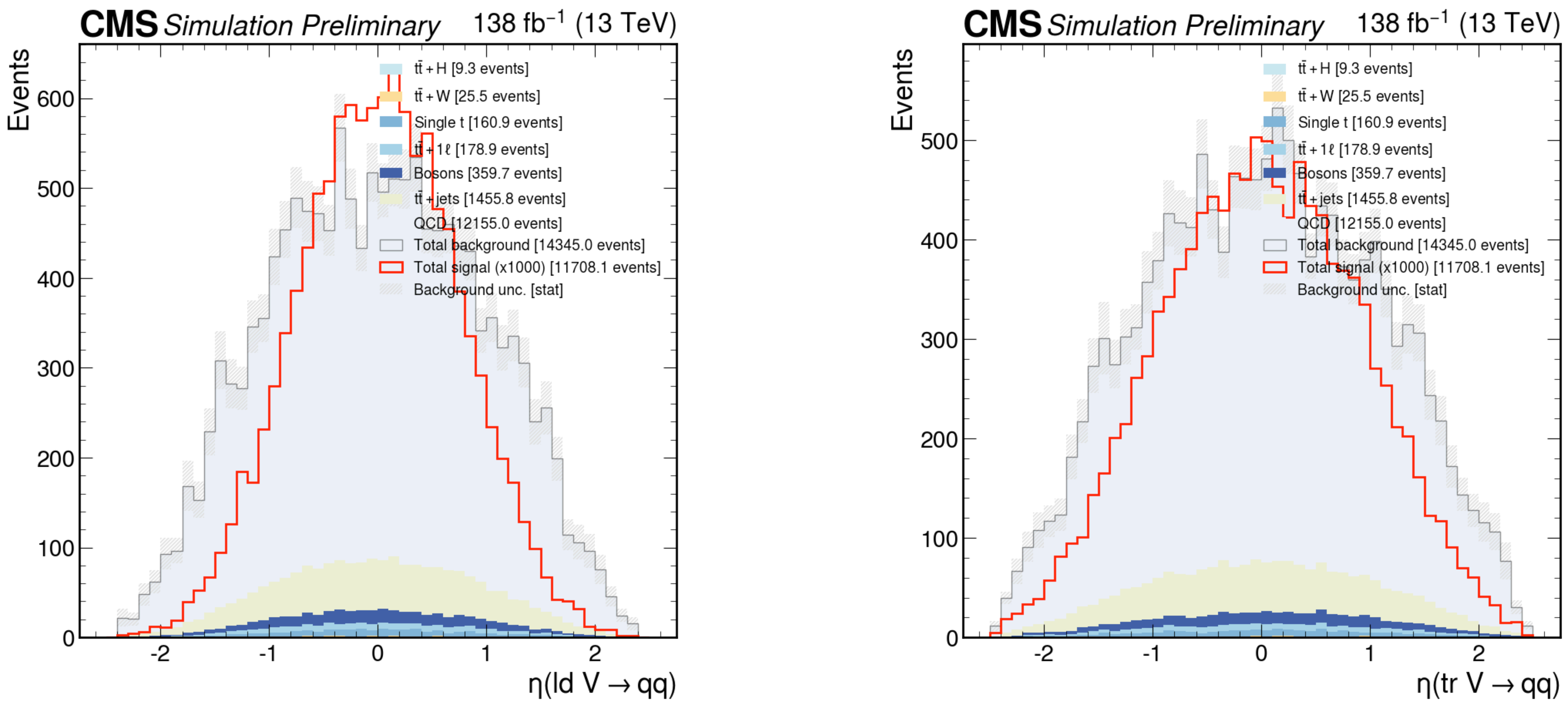
Cut	QCD	$t\bar{t}$+jets	$t\bar{t}+1\ell$	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	18,030K	157K	26K	979	430	13K	344K	18,570K	—	133	—
HLT + MET Filters	17,942K	156K	26K	975	428	13K	341K	18,479K	100%	132	99%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	2%	32	24%
Object selection	158K	6.2K	855	59	30	478	5.1K	171K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%
ParticleNet score cuts	949	226	28	6	2	47	55	1.3K	9%	8	65%
$\|\Delta\eta_{jj}\| > 4$	215	48	6	1	0	13	10	294	22%	7	85%
Signal Region (ABCDNet > 0.96)	0.67	0.07	0.14	0.00	0.00	0.15	0.43	1.46	0%	3.73	57%

Analysis Handles: H→bb (Preselection)



η distribution is slightly more narrow for signal

Analysis Handles: V→qq (Preselection)



η distribution is slightly more narrow for signal

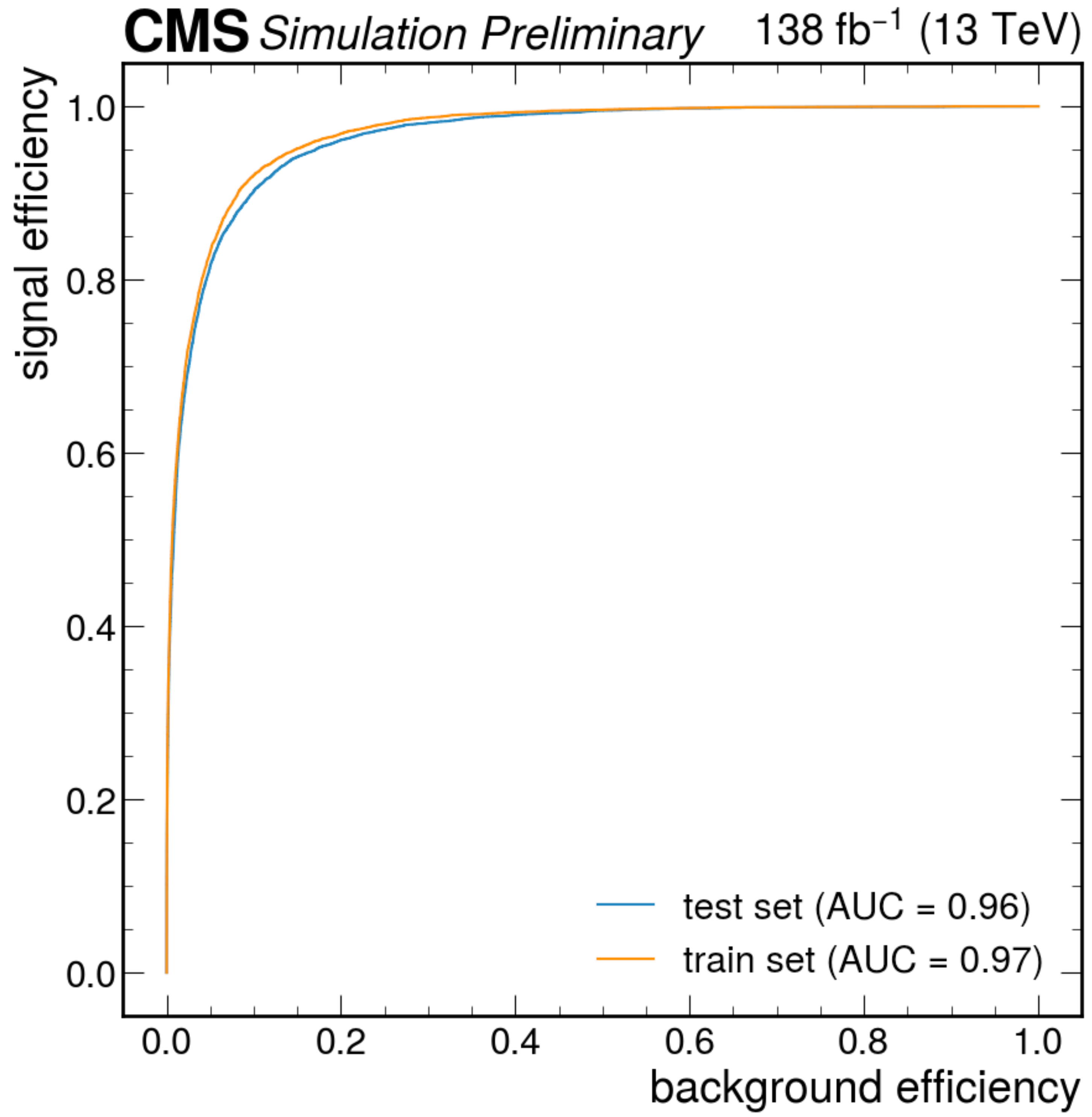
BDT-based Signal Region

NOT USED

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}$+1ℓ	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	137,061K	748K	86K	2.6K	1.3K	53K	1,513K	139,464K	—	175	—
HLT + MET Filters	88,702K	575K	70K	2.2K	1.1K	41K	1,120K	90,512K	65%	168	96%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	0%	32	19%
Object selection	160K	6.2K	862	59	30	482	5.2K	172K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%

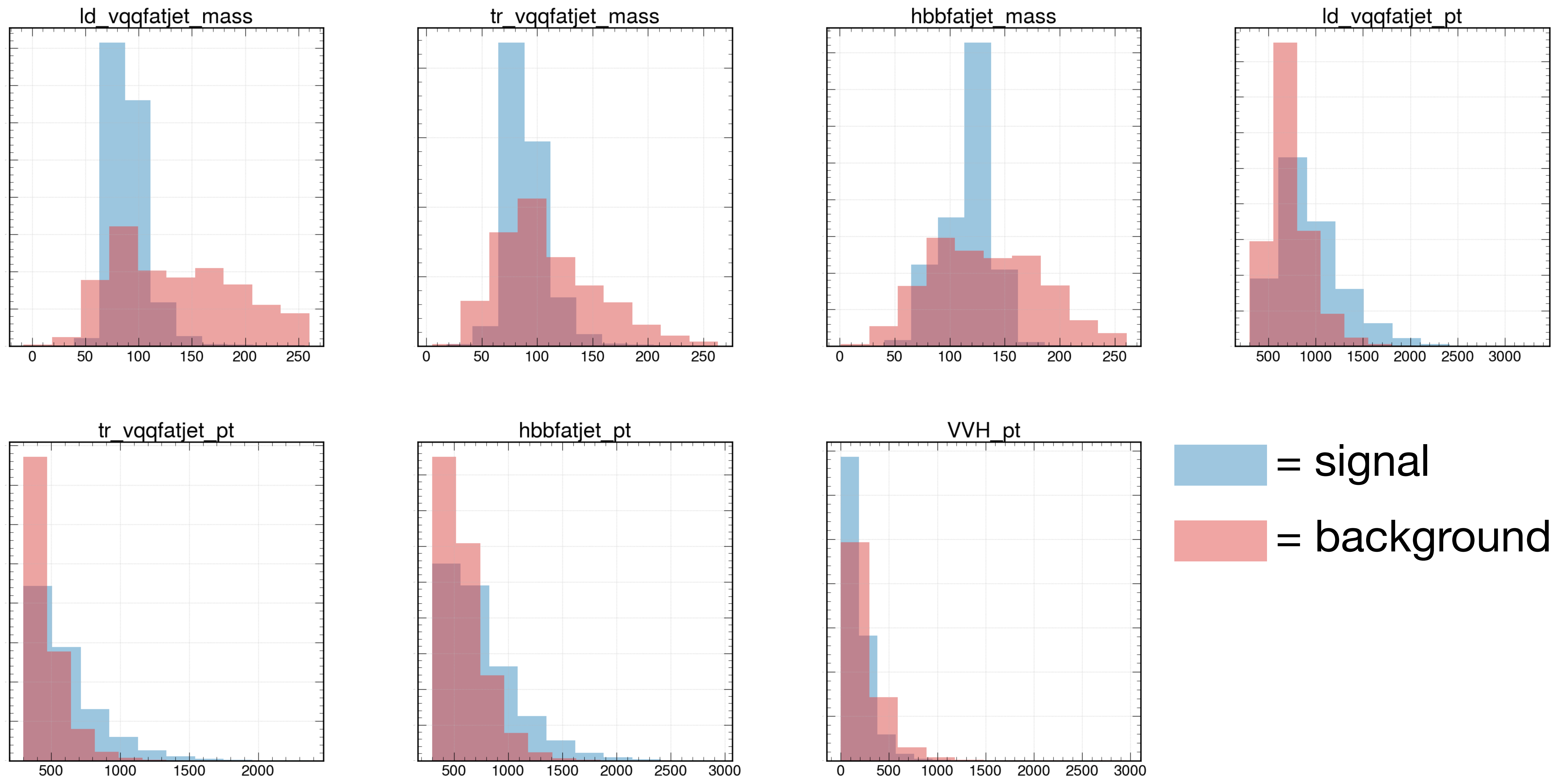
- Train a simple BDT to get a “ceiling” for analysis sensitivity
- Use the following inputs:
  - H→bb fat jet p4 (p_T, η, ϕ), M_{PNet}
  - V→qq fat jet p4 (p_T, η, ϕ), M_{PNet}
  - VVH system p_T
- BDT hyperparameters tabulated in backup



UC San Diego

NOT USED

BDT



Using matplotlib's weird automatic binning
(for visualization purposes only)

Parameter	Value	Description*
`objective`	binary:logistic	Learning objective; 'binary:logistic' specifies logistic regression for binary classification, output probability
`eta`	0.1	Step size shrinkage (alias: learning_rate)
`max_depth`	3	Max. depth of tree: larger = more complex = more prone to overfitting
`verbosity`	1	0 (silent), 1 (warning), 2 (info), 3 (debug)
`nthread`	8	Number of parallel threads
`eval_metric`	auc	Evaluation metrics for validation data. 'auc' = Area Under the Curve
`subsample`	0.6	Subsample ratio of the training instances
`alpha`	8.0	L1 regularization term on weights: Larger = more conservative
`gamma`	2.0	Min. loss rediction to make leaf (alias: min_split_loss)
`lambda`	1.0	L2 regularization term on weights: Larger = more conservative
`min_child_weight`	1.0	Minimum sum of instance weight (hessian) needed in a child
`colsample_bytree`	1.0	The subsample ratio of columns when constructing each tree
`scale_pos_weight`	2456.3	Control the balance of positive and negative weights, useful for unbalanced classes

*From: https://xgboost.readthedocs.io/en/stable/parameter.html

UC San Diego

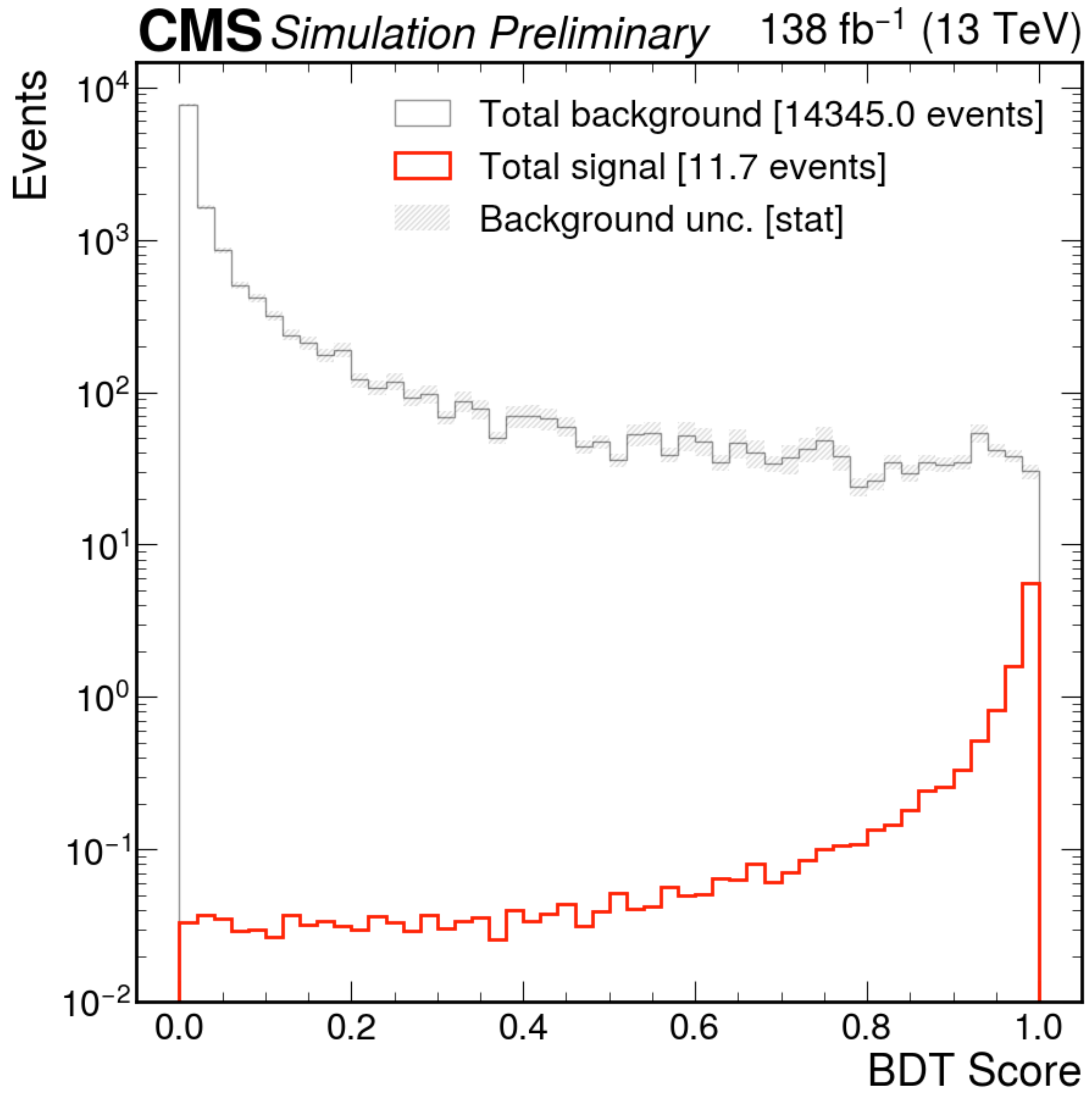
NOT USED

BDT-based Signal Region

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}+1\ell$	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH ($C_{2V} = 2$)	Eff.
Skim	137,061K	748K	86K	2.6K	1.3K	53K	1,513K	139,464K	—	175	—
HLT + MET Filters	88,702K	575K	70K	2.2K	1.1K	41K	1,120K	90,512K	65%	168	96%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	0%	32	19%
Object selection	160K	6.2K	862	59	30	482	5.2K	172K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%

- Brute-force scan over the following cuts:
  - BDT score
  - ParticleNet scores (Xbb, XWqq)
  - M_{jj}
  - $|\Delta\eta_{jj}|$
- Optimize for $S/\sqrt{B}$ as a significance heuristic



UC San Diego

NOT USED

BDT-based Signal Region

Yields scaled to lumi×σ, rounded for readability

Cut	QCD	$t\bar{t}$+jets	$t\bar{t}+1\ell$	$t\bar{t}$+W	$t\bar{t}$+H	Single top	Bosons	Total Bkg.	Eff.	VBSVVH (C_{2V} = 2)	Eff.
Skim	137,061K	748K	86K	2.6K	1.3K	53K	1,513K	139,464K	—	175	—
HLT + MET Filters	88,702K	575K	70K	2.2K	1.1K	41K	1,120K	90,512K	65%	168	96%
At least 3 fat jets	395K	9.8K	1.4K	110	46	874	13K	421K	0%	32	19%
Object selection	160K	6.2K	862	59	30	482	5.2K	172K	41%	18	56%
Preselection	12K	1.5K	179	25	9	161	360	14K	8%	12	66%
Signal Region	0.14	0.37	0.04	-0.02	0.00	0.08	0.18	0.81	0%	5	43%

- Settled on the following signal region:
  BDT > 0.9 and Xbb > 0.5 and XWqq > 0.82|0.66 (ld|tr) and $|\Delta\eta_{jj}| > 4$ and $M_{jj} > 600$ GeV
- Therefore the "ceiling" is high: **5 signal vs. 1 background for C_{2V} = 2**
- Next step: background extrapolation
  - We would like it to be data-driven, as we do not trust QCD
  - **Spoiler:** ABCD does not work with the BDT, so we pivot to a novel technique

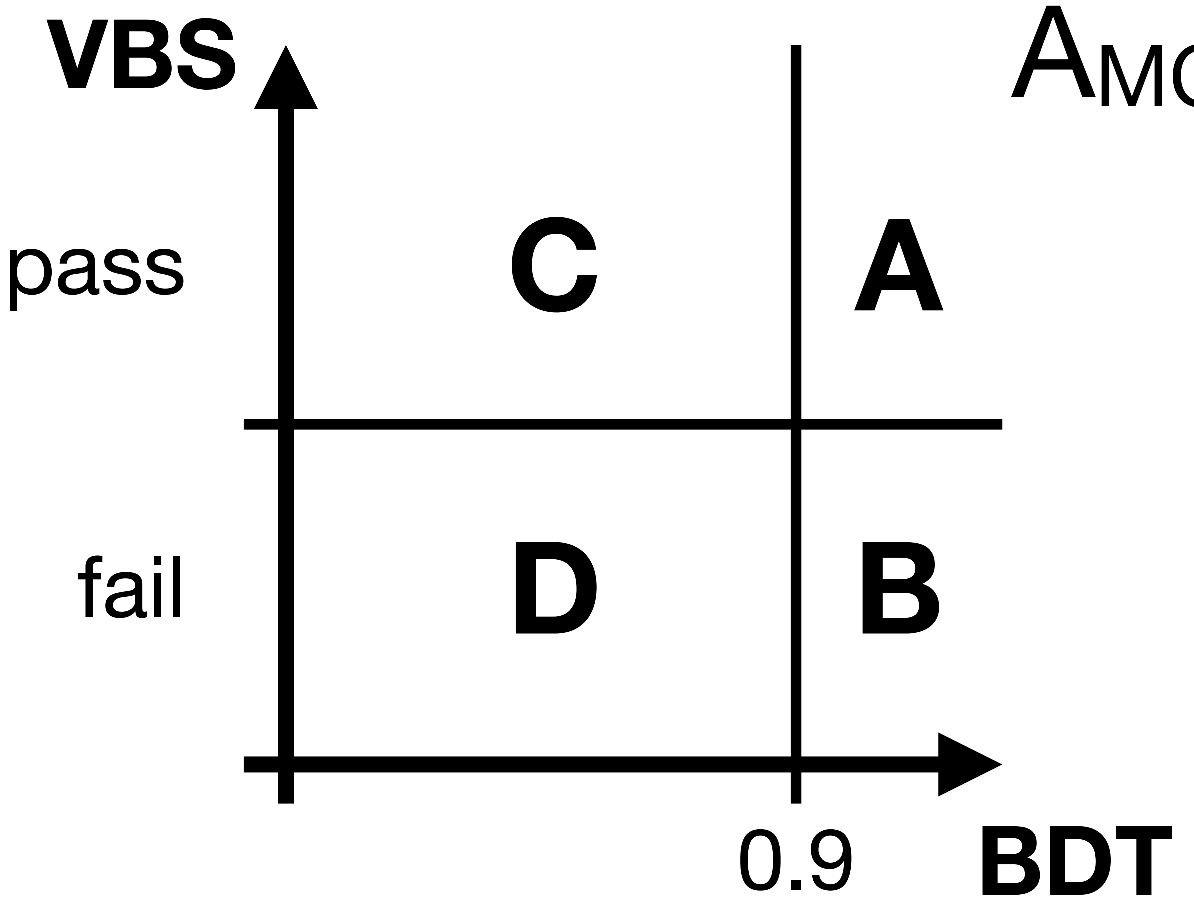
NOT USED

ABCD: BDT Signal Region

Xbb > 0.50 and XWqq > 0.82|0.66 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
$\lvert\Delta\eta_{jj}\rvert > 4$ and $M_{jj} > 600$ GeV and BDT > 0.9	A	0.81	0.28	5.05	0.06	—	—
$\lvert\Delta\eta_{jj}\rvert \leq 4$ and $M_{jj} \leq 600$ GeV and BDT > 0.9	B	20.87	2.62	1.03	0.03	25	5.00
$\lvert\Delta\eta_{jj}\rvert > 4$ and $M_{jj} > 600$ GeV and BDT ≤ 0.9	C	175.92	20.87	1.15	0.03	172	13.11
$\lvert\Delta\eta_{jj}\rvert \leq 4$ and $M_{jj} \leq 600$ GeV and BDT ≤ 0.9	D	1069.99	44.43	0.32	0.02	1190	34.50

Very performant SR, but terrible MC closure



$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = 3.43 \pm 0.61$$

$$\epsilon_{syst} = \left|1 - \frac{A^{pred}_{MC}}{A_{MC}}\right| = \left|1 - \frac{3.43}{0.81}\right| = 325.0\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 21.6\%$$

Final Result

Expected sig. 5.05 ± 0.06

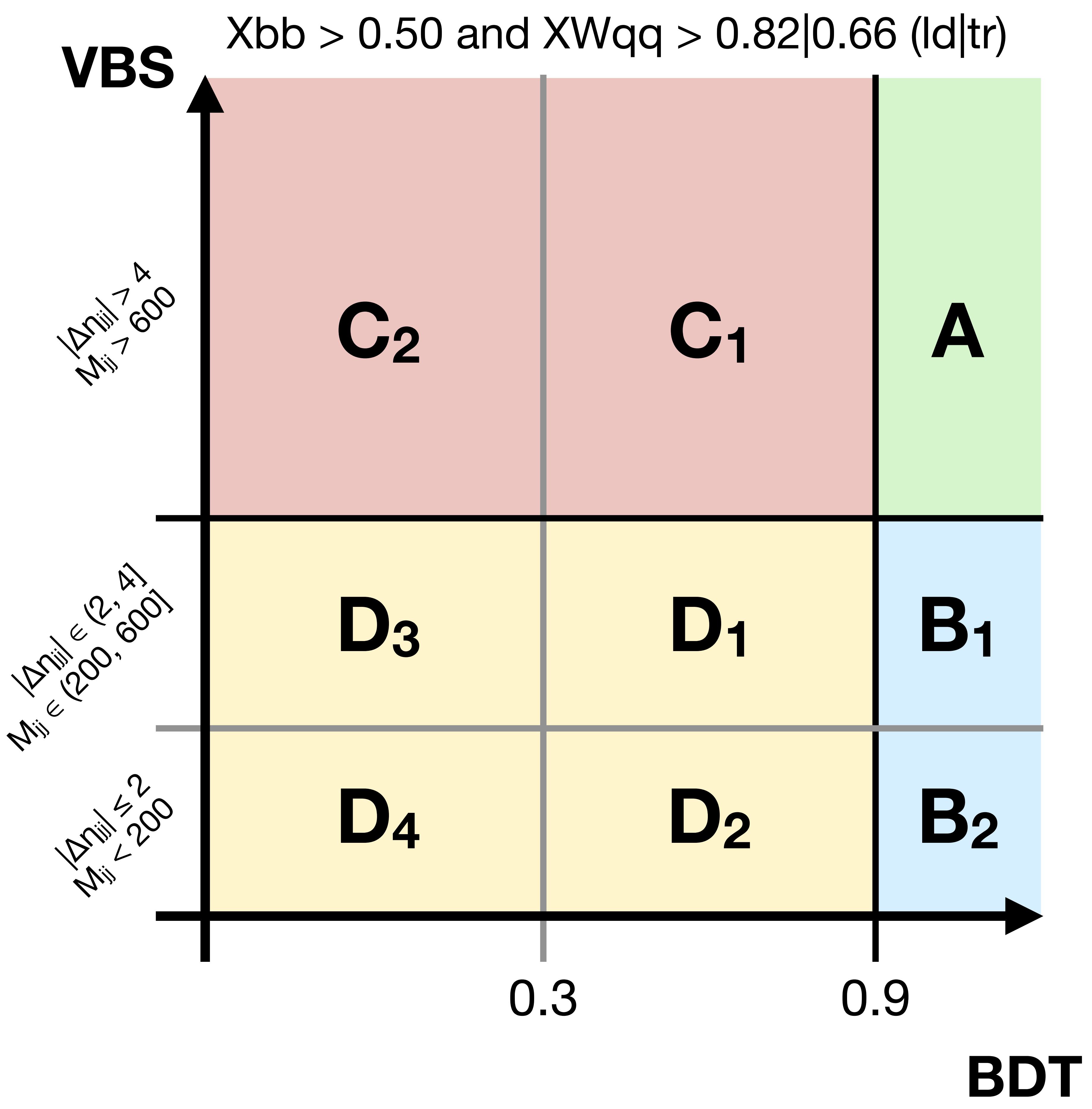
Predicted bkg. 3.61 ± 0.78 ± 11.7

stat. *syst.*

Terrible closure ⇒ BDT & VBS correlated or poor MC composition modeling?

UC San Diego

BDT



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.81**	0.28	**5.05**	0.06	–	–
B	20.87	2.62	1.03	0.03	25	5.00
C	175.92	20.87	1.15	0.03	172	13.11
D	1070.0	44.43	0.32	0.02	1190	34.50

$$A^{\text{pred}} = B \times \frac{C}{D} = \left[\begin{array}{l} 3.43 \pm 0.61 \ \ \text{(MC)} \\ \mathbf{3.61 \pm 0.78} \ \ \text{(Data)} \end{array}\right.$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	12.42	1.85	0.59	0.02	**16**	4.00
B_2	8.44	1.86	0.44	0.02	9	3.00
D_1	93.31	12.75	0.15	0.01	87	9.33
D_2	43.38	4.16	0.09	0.01	55	7.42

$$B_1^{\text{pred}} = B_2 \times \frac{D_1}{D_2} = \mathbf{14.24 \pm 5.34} \ \text{(Data)} \ ✓$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	93.31	12.75	0.15	0.01	**87**	9.33
D_2	43.38	4.16	0.09	0.01	55	7.42
D_3	604.79	34.56	0.04	0.01	649	25.48
D_4	328.51	24.49	0.03	0.00	399	19.97

$$D_1^{\text{pred}} = D_2 \times \frac{D_3}{D_4} = \mathbf{89.46 \pm 13.3} \ \text{(Data)} \ ✓$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	19.72	7.16	1.01	0.03	**19**	4.36
D_1	93.31	12.75	0.15	0.01	87	9.33
C_2	156.20	19.61	0.13	0.01	153	12.37
D_3	604.79	34.56	0.04	0.01	649	25.48

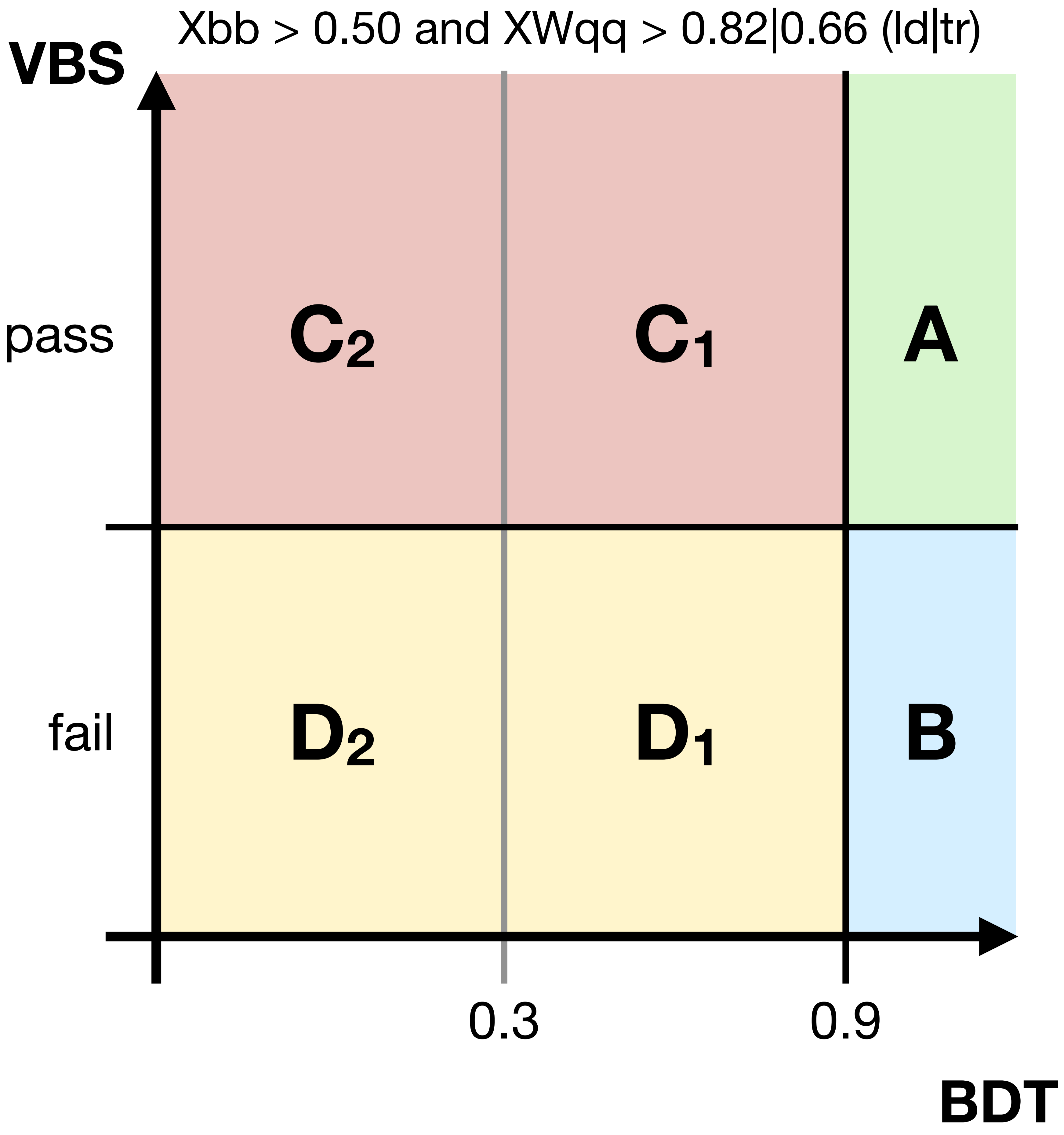
$$C_1^{\text{pred}} = D_1 \times \frac{C_2}{D_3} = \mathbf{20.51 \pm 2.87} \ \text{(Data)} \ ✓$$

ABCD works well in data (in sidebands), but predicted S/√B is not good

Studied for some time, then pivoted to "Automated ABCD" (next slides)

UC San Diego

BDT



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.81**	0.28	**5.05**	0.06	–	–
B	20.87	2.62	1.03	0.03	25	5.00
C	175.92	20.87	1.15	0.03	172	13.11
D	1070.0	44.43	0.32	0.02	1190	34.50

$$A^{pred} = B \times \frac{C}{D} = \left[\begin{array}{l} 3.43 \pm 0.61 \ \text{(MC)} \\ \mathbf{3.61 \pm 0.78} \ \text{(Data)} \end{array} \right.$$

✓

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.81**	0.28	**5.05**	0.06	–	–
B	20.87	2.62	1.03	0.03	25	5.00
C_1	19.72	7.16	1.01	0.03	19	4.36
D_1	136.7	13.42	0.25	0.01	142	11.92

$$A^{pred} = B \times \frac{C_1}{D_1} = \left[\begin{array}{l} 3.01 \pm 1.19 \ \text{(MC)} \\ \mathbf{3.35 \pm 1.06} \ \text{(Data)} \end{array} \right.$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	19.72	7.16	1.01	0.03	**19**	4.36
D_1	136.69	13.42	0.25	0.01	142	11.92
C_2	156.20	19.61	0.13	0.01	153	12.37
D_2	933.30	42.35	0.07	0.01	1048	32.37

$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{20.73 \pm 2.50} \ \text{(Data)} \ ✓$$

ABCD works well in sidebands, but predicted S/√B is worse than ABCDNet

UC San Diego

Backup: MVA & Bkg. Estimation

Automated ABCD

- Introduce Distance Correlation
  - DisCo for catchy titles, dCorr for math
  - dCorr(f, g) = 0 iff f and g are independent
  - dCorr(f, g) ∈ (0, 1] otherwise
- Claim high performance/easy to train vs. other decorrelation metrics
- Added to some typical loss (assume BCE)
- Hyperparameter λ controls relative size of DisCo term vs. BCE

PHYSICAL REVIEW D **103**, 035021 (2021)

Automating the ABCD method with machine learning

Gregor Kasieczka,[1,*] Benjamin Nachman,[2,†] Matthew D. Schwartz,[3,§] and David Shih[2,4,5,‡]

[1]*Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, D-22761 Hamburg, Germany*
[2]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[3]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[4]*NHETC, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*
[5]*Berkeley Center for Theoretical Physics, University of California, Berkeley, California 94720, USA*

(Received 6 August 2020; accepted 3 February 2021; published 22 February 2021)

$$\mathcal{L}[f(X)] = \mathcal{L}_{\text{classifier}}[f(X), y] + \lambda \mathrm{dCorr}^2_{y=0}[f(X), X_0], \quad (3.1)$$

$$\mathcal{L}[f, g] = \mathcal{L}_{\text{classifier}}[f(X), y] + \mathcal{L}_{\text{classifier}}[g(X), y] + \lambda \mathrm{dCorr}^2_{y=0}[f(X), g(X)], \quad (3.2)$$

APPENDIX A: DISTANCE CORRELATION

For two random variables f and g, the distance covariance is defined as

$$\begin{aligned} \mathrm{dCov}^2[f,g] = {} & \langle |f - f'| \times |g - g'| \rangle \\ & + \langle |f - f'| \rangle \times \langle |g - g'| \rangle \\ & - 2\langle |f - f'| \times |g - g''| \rangle, \end{aligned} \quad \text{(A1)}$$

where (f, g), (f', g'), (f'', g'') are all independent and identically distributed from the same joint distribution. In practice, we evaluate $\mathrm{dCov}^2[f, g]$ by averaging $|f_i - f_j| \times |g_i - g_j|$, $|f_i - f_j|$, and $|g_i - g_j|$ over all pairs of events i, j, and $|f_i - f_j| \times |g_i - g_k|$ over all triplets of events i, j, k.

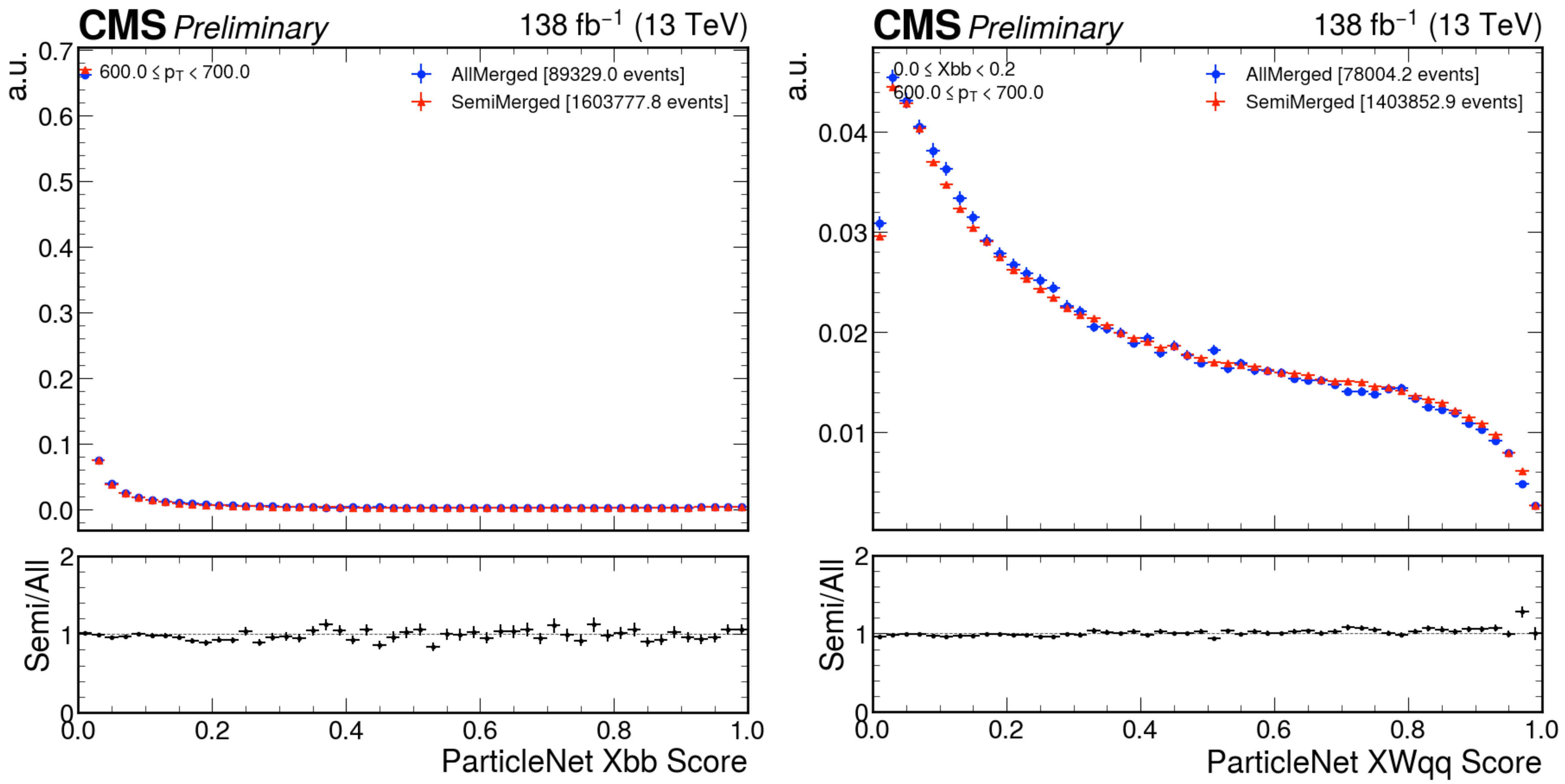
The distance correlation is then defined analogously to the usual correlation:

$$\mathrm{dCorr}^2[f,g] = \frac{\mathrm{dCov}^2[f,g]}{\mathrm{dCov}[f,f]\mathrm{dCov}[g,g]}. \quad \text{(A2)}$$

CMS

UC San Diego

QCD Corrections

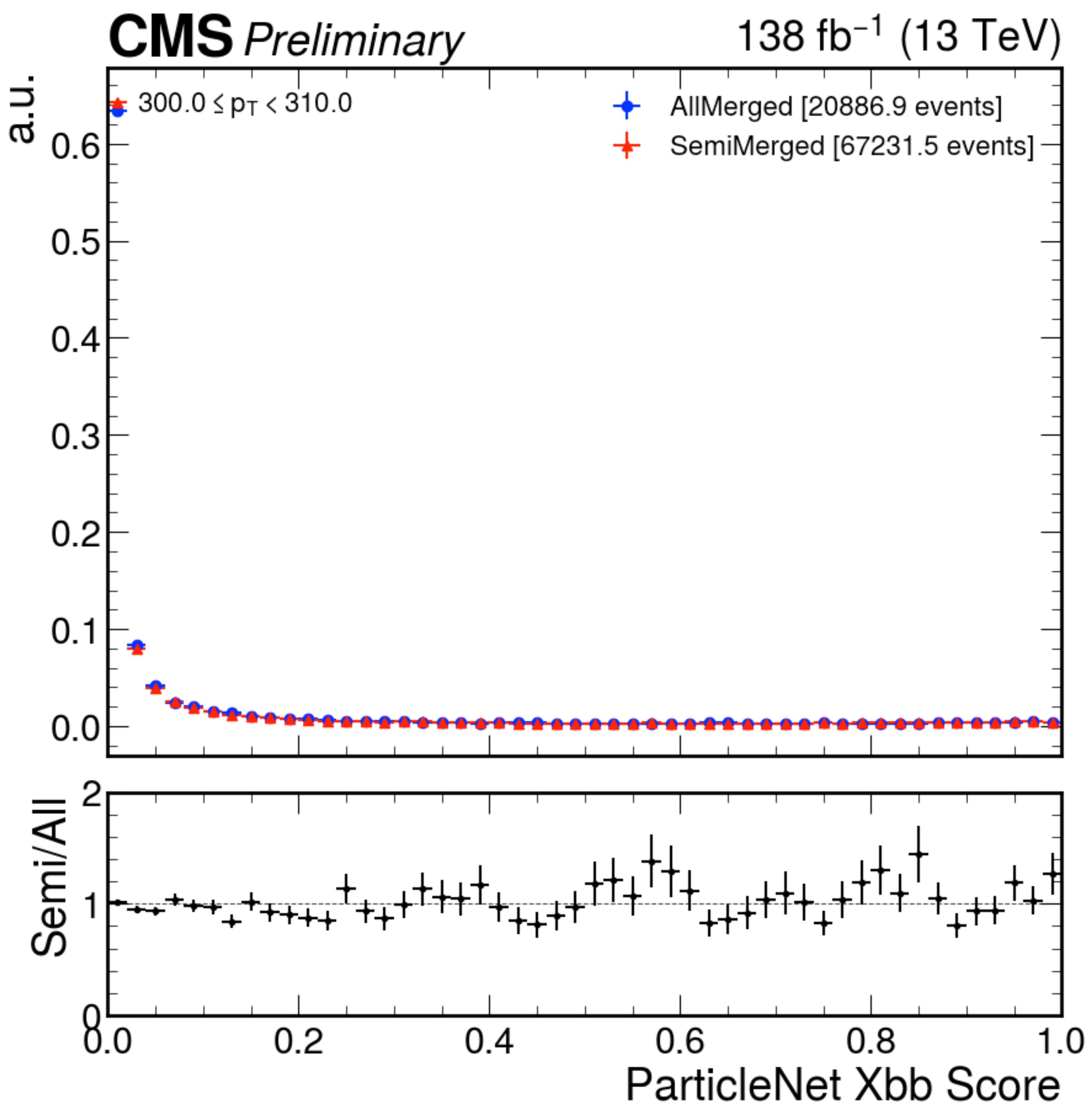


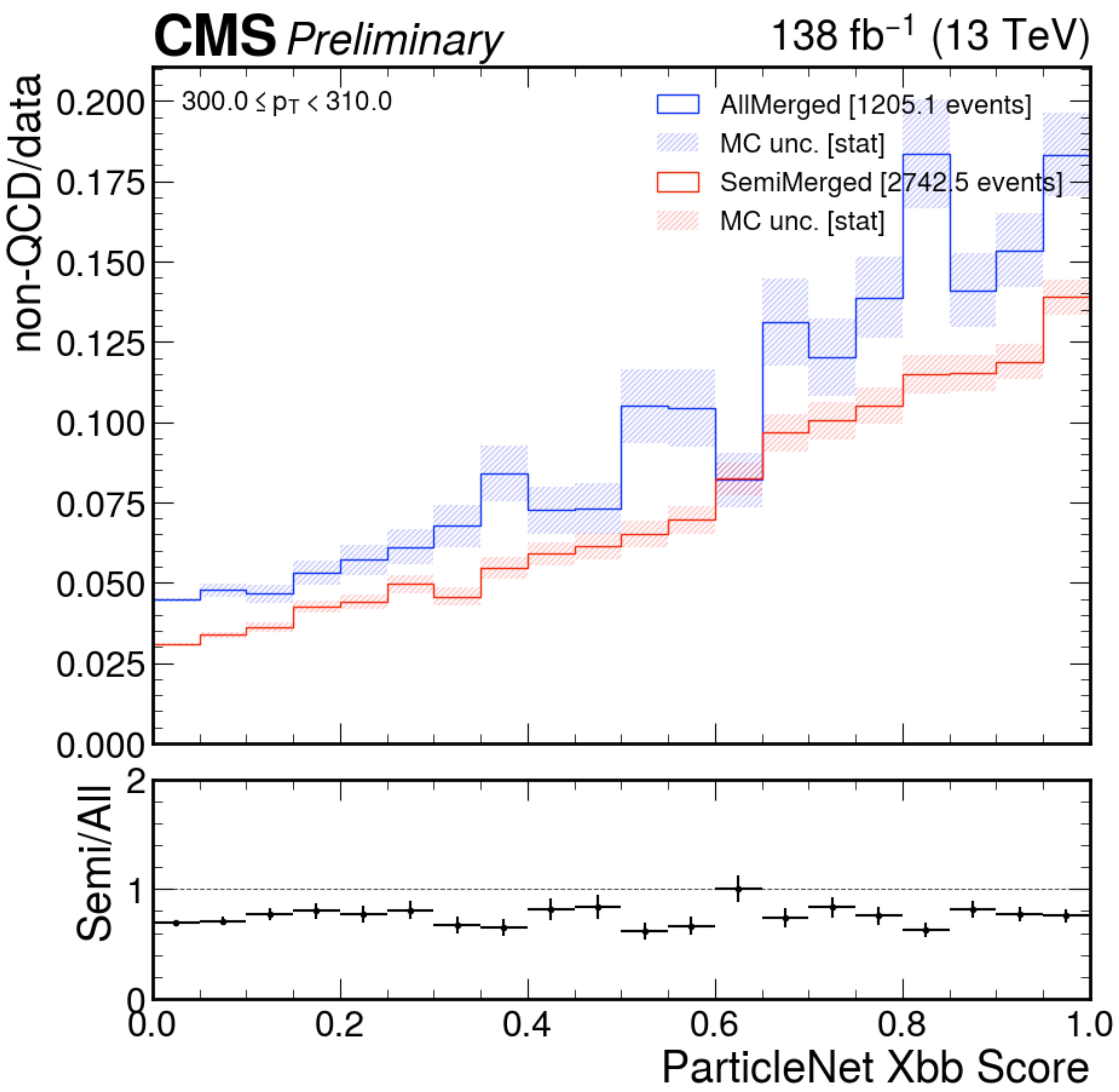
- All plots can be found here
- ParticleNet scores are replaced:

```
for fatjet in fatjets:
    // Sample PDFs for Xbb and XWqq scores
    pt = fatjet.pt()
    xbb = xbb_pdf2D.ProjectionY(pt).GetRandom()
    xwqq = xwqq_pdf3D.ProjectionZ(pt, xbb).GetRandom()
    // Replace Xbb and XWqq scores
    fatjet.xbb = xbb
    fatjet.xwqq = xwqq
```

- **The "projection" methods are pseudocode shorthand** for getting a slice of an N-dimensional histogram along only one axis in one bin of the others
- XWqq PDF is binned in p_T and Xbb score because $H \to b\bar{b}$ candidate is selected by max(Xbb) before the $V \to qq$ candidates are selected

QCD Corrections: Non-QCD Fraction

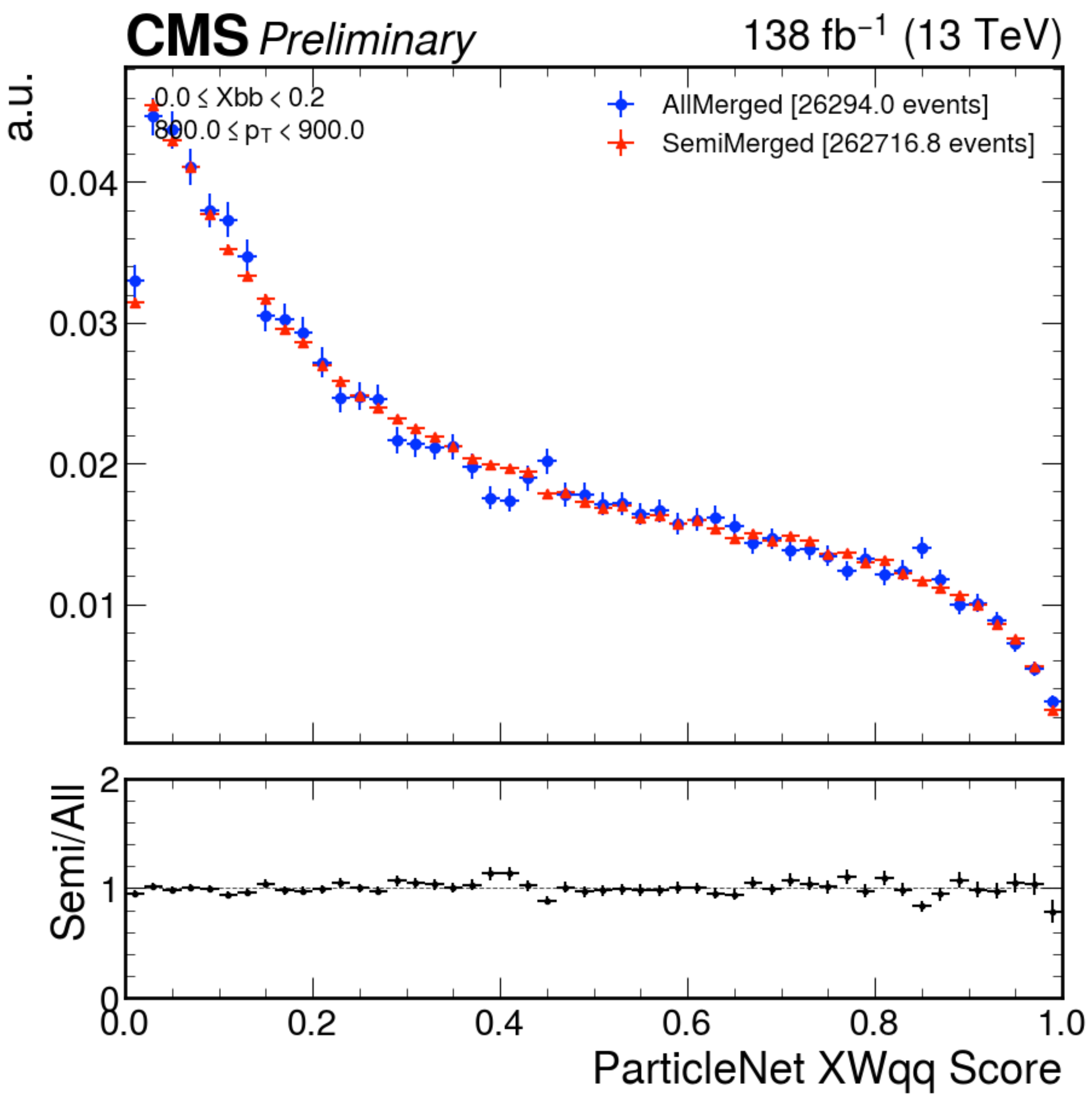


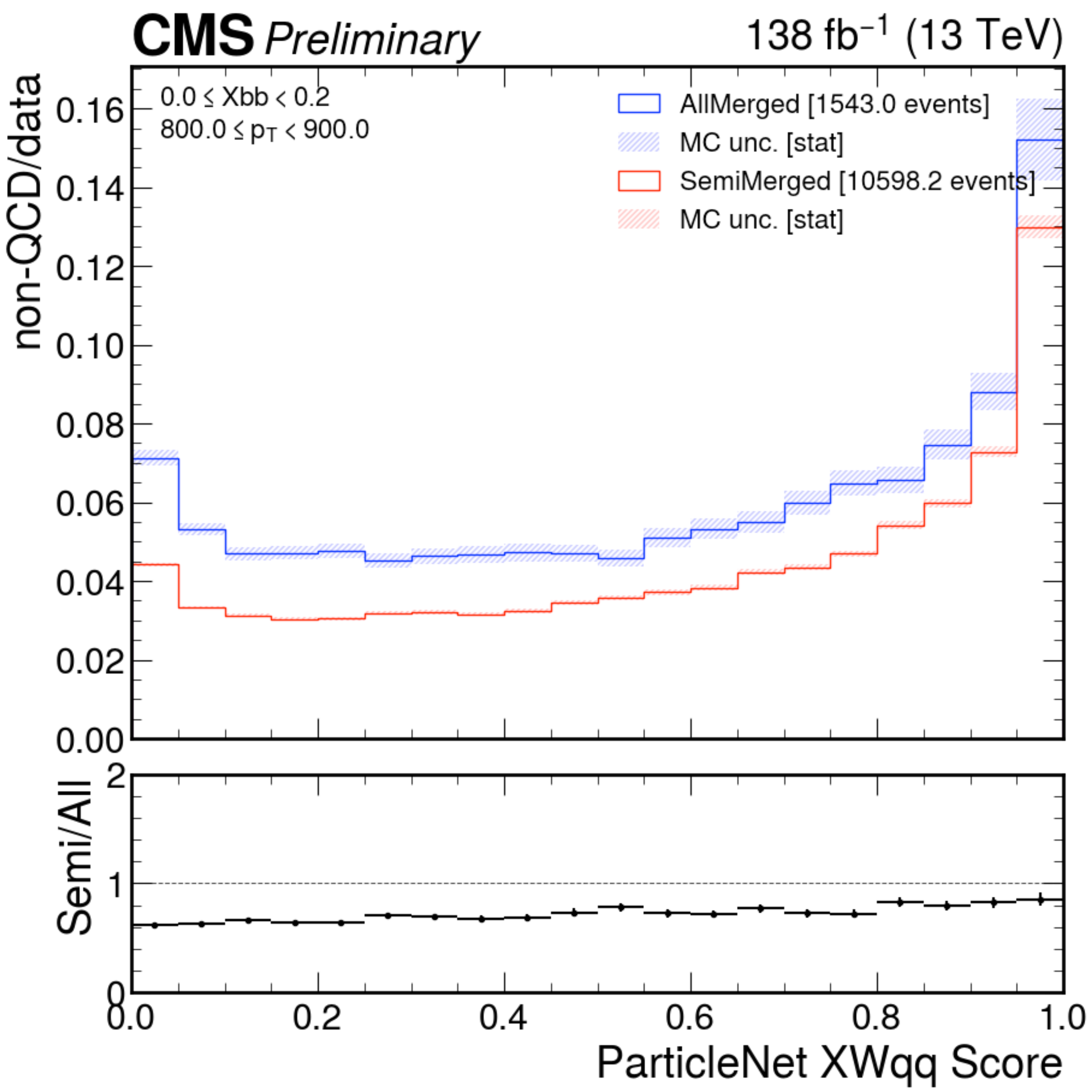


non-QCD is a small fraction of total (never exceeds 25%)

http://uaf-10.t2.ucsd.edu/~jguiang/vbsvvhjets_plots/qcdcorr_xbbbins/#data_minus_nonqcd_xbb
http://uaf-10.t2.ucsd.edu/~jguiang/vbsvvhjets_plots/qcdcorr_xbbbins/#nonqcd_frac_xbb

QCD Corrections: Non-QCD Fraction



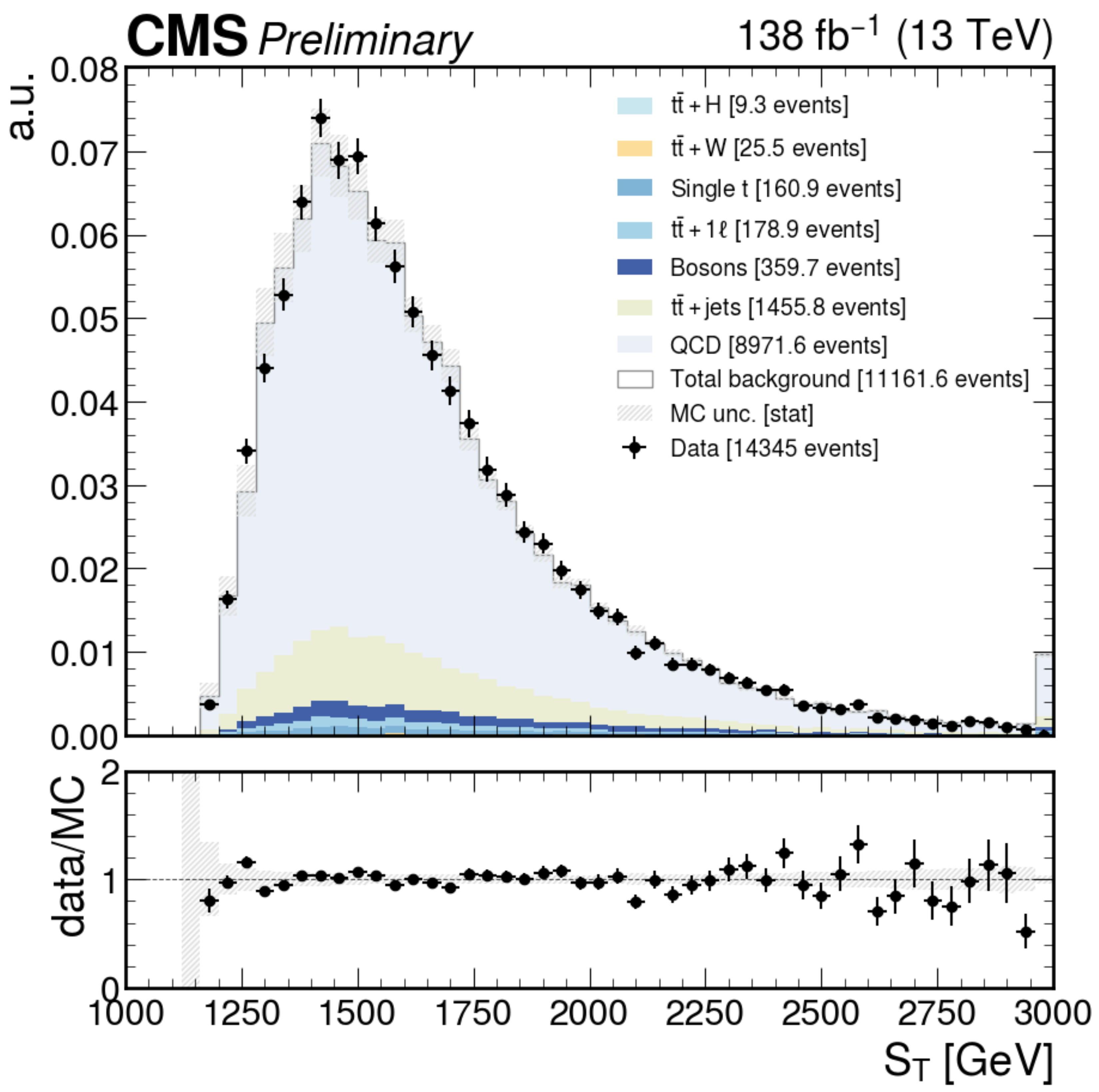


non-QCD is a small fraction of total (never exceeds 25%)

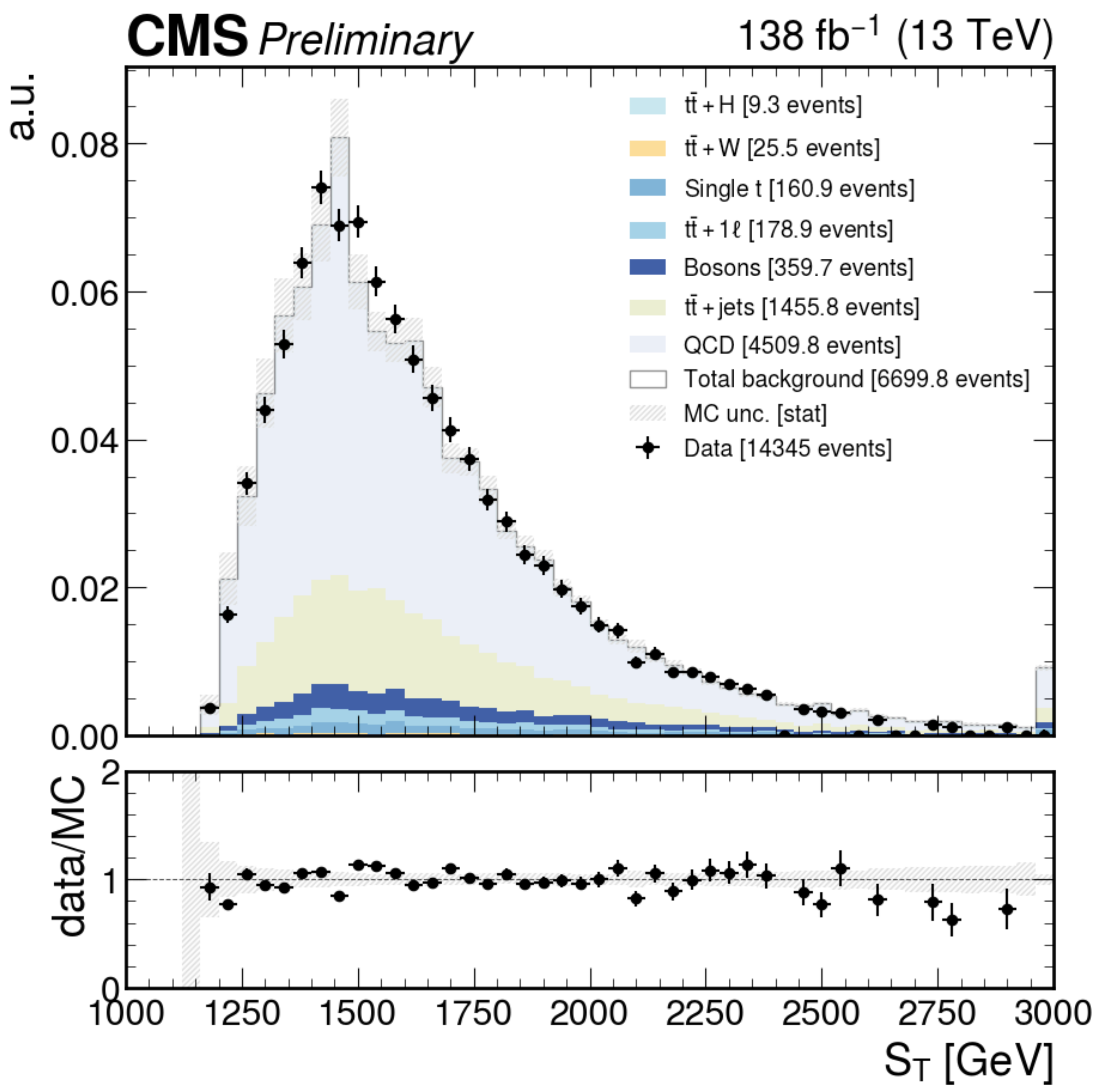
http://uaf-10.t2.ucsd.edu/~jguiang/vbsvvhjets_plots/qcdcorr_xbbbins/#data_minus_nonqcd_xwqq
http://uaf-10.t2.ucsd.edu/~jguiang/vbsvvhjets_plots/qcdcorr_xbbbins/#nonqcd_frac_xwqq

QCD Corrections: With vs. Without

With



Without

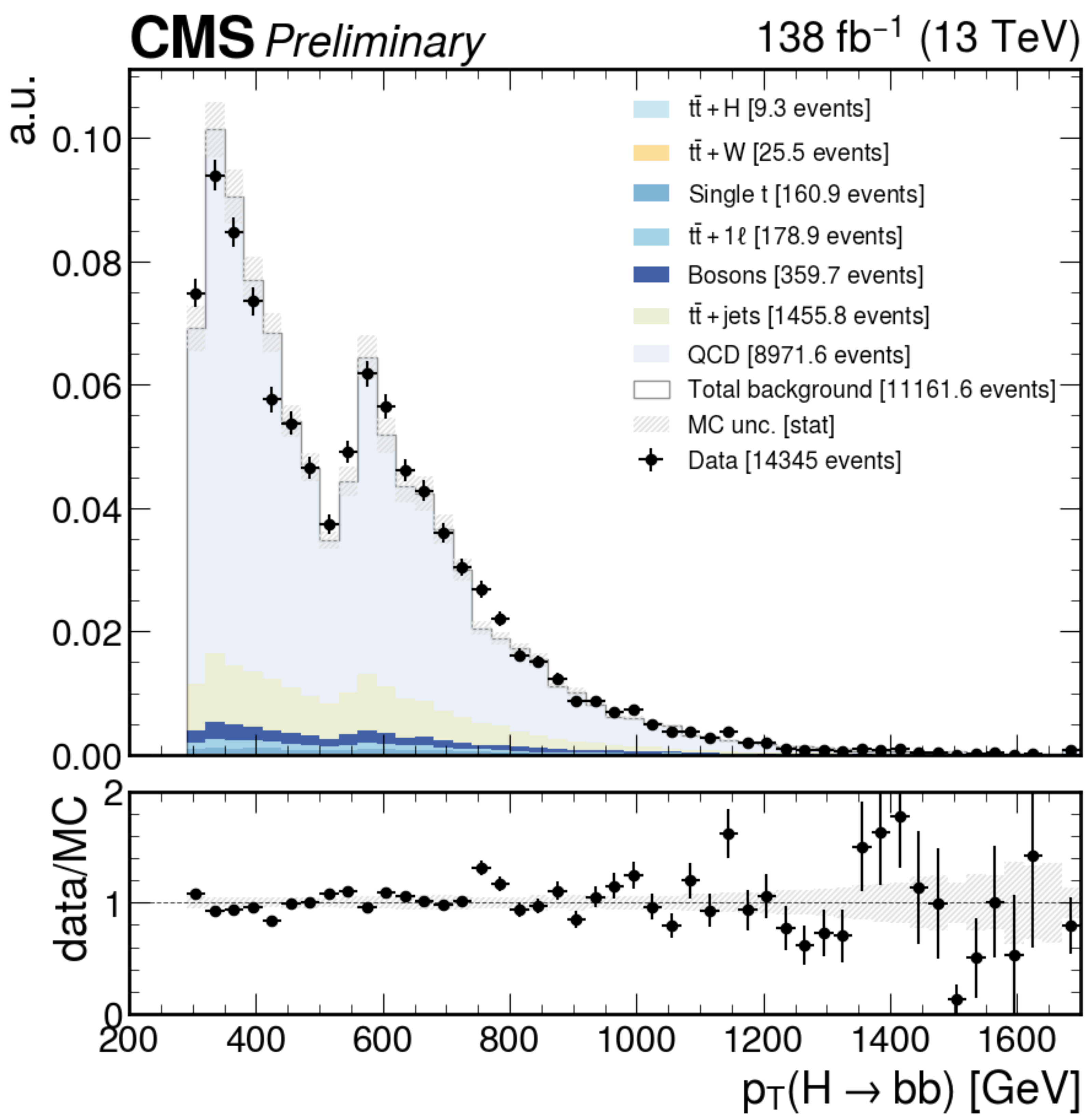


About 2x more weighted QCD events with correction after Preselection

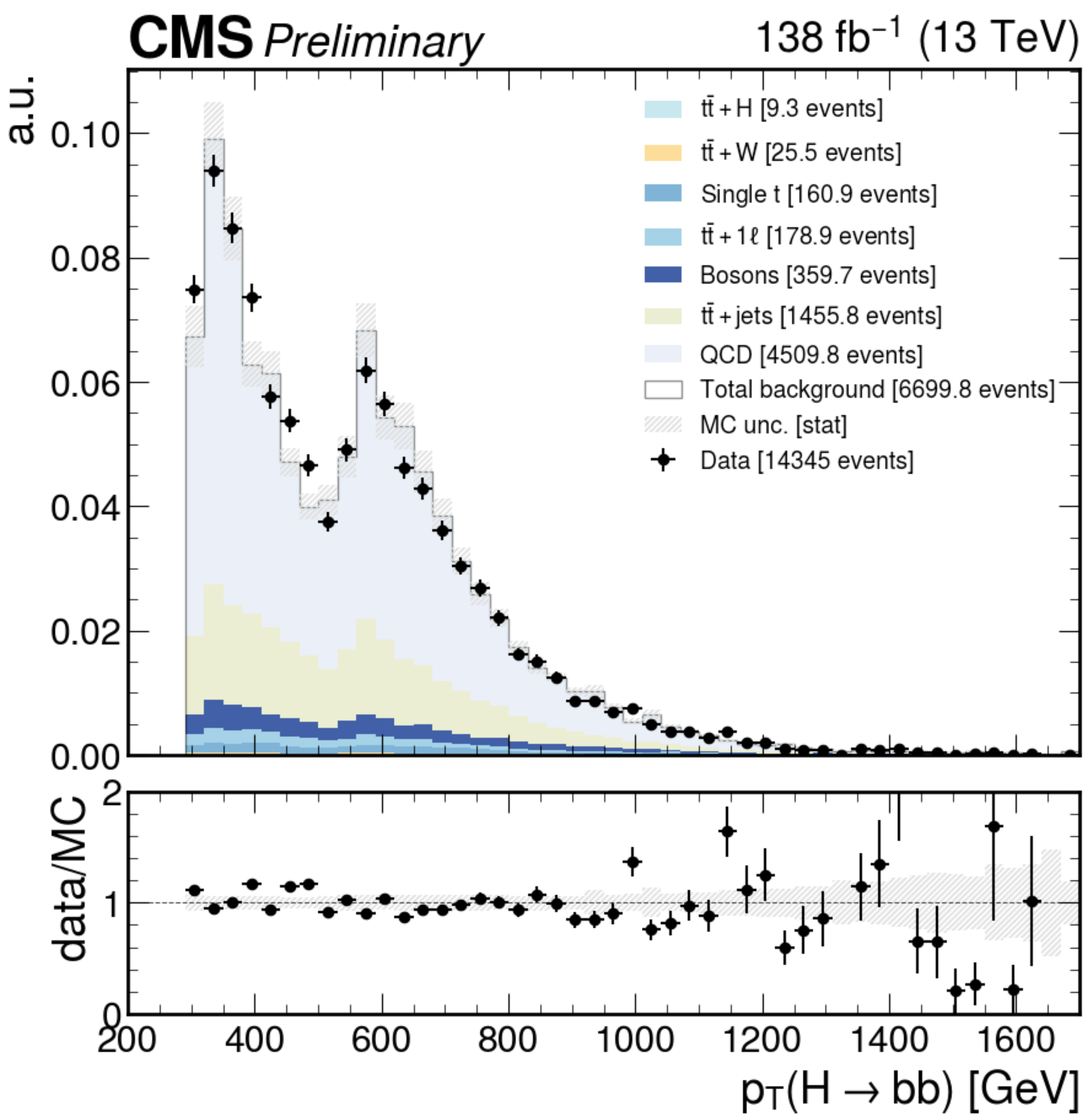
QCD integral is normalized to (data - non-QCD) after this step

QCD Corrections: With vs. Without

With



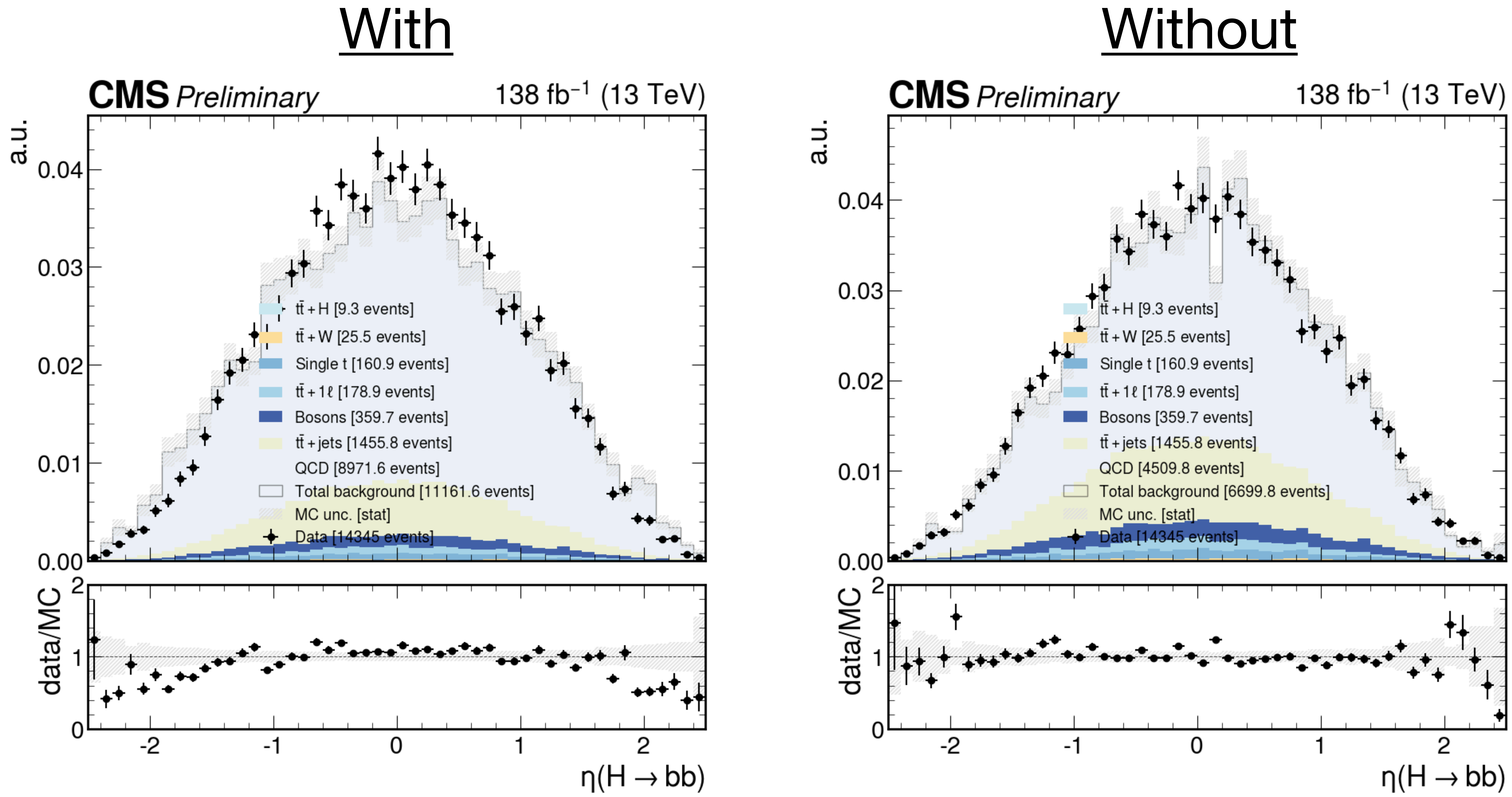
Without



About 2x more weighted QCD events with correction after Preselection

QCD integral is normalized to (data - non-QCD) after this step

QCD Corrections: With vs. Without

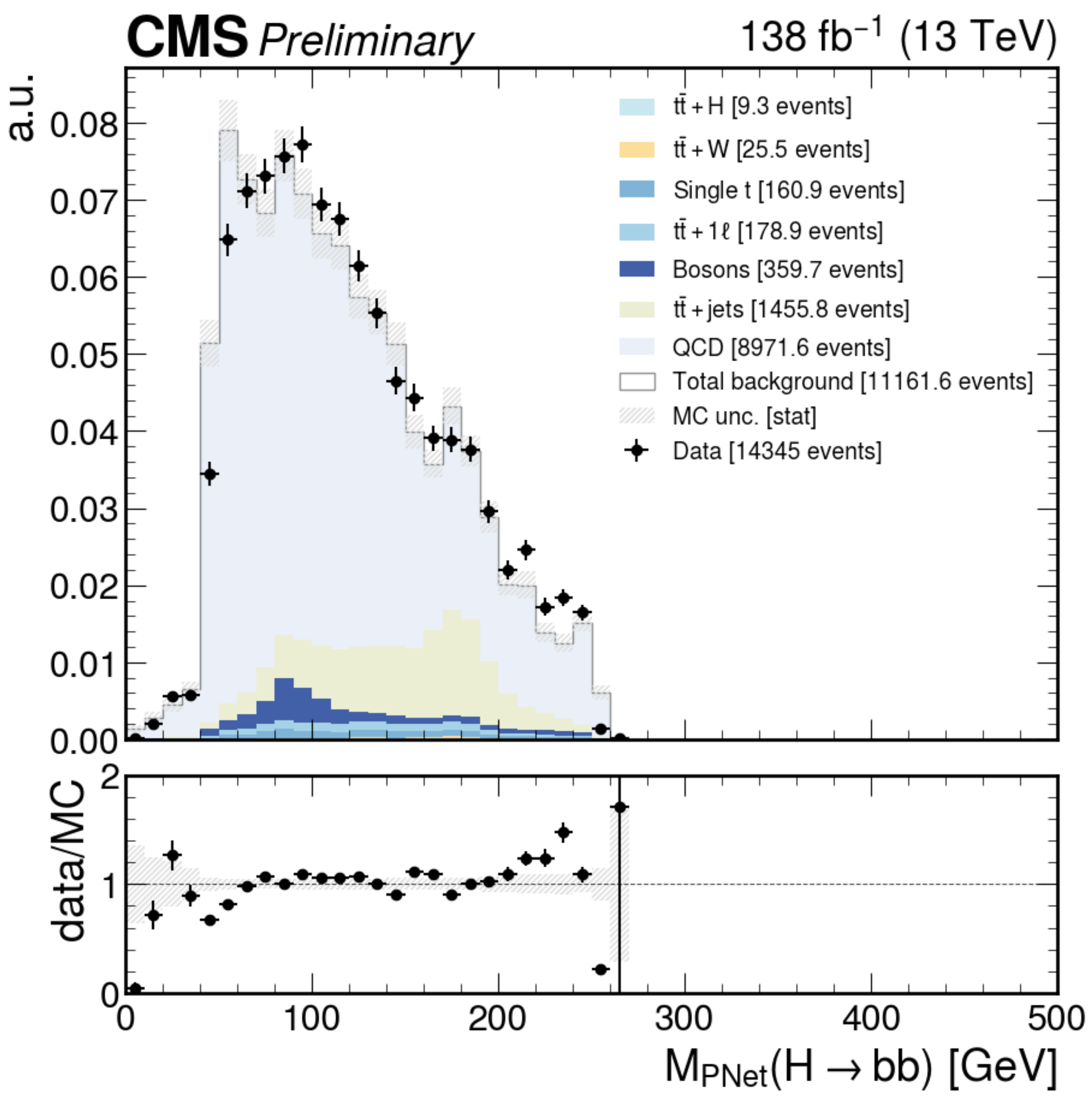


About 2x more weighted QCD events with correction after Preselection

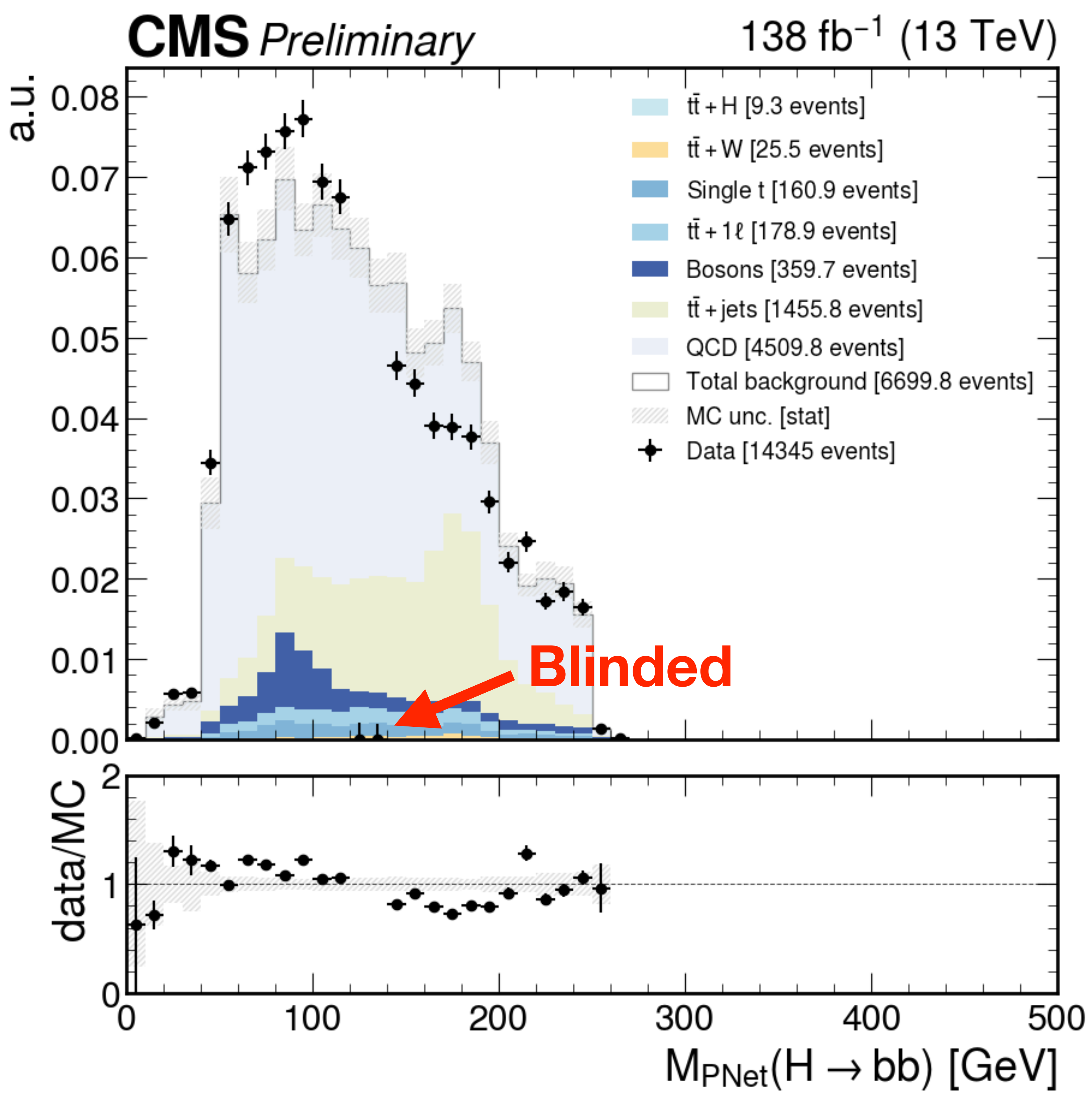
QCD integral is normalized to (data - non-QCD) after this step

QCD Corrections: With vs. Without

<u>With</u>



<u>Without</u>

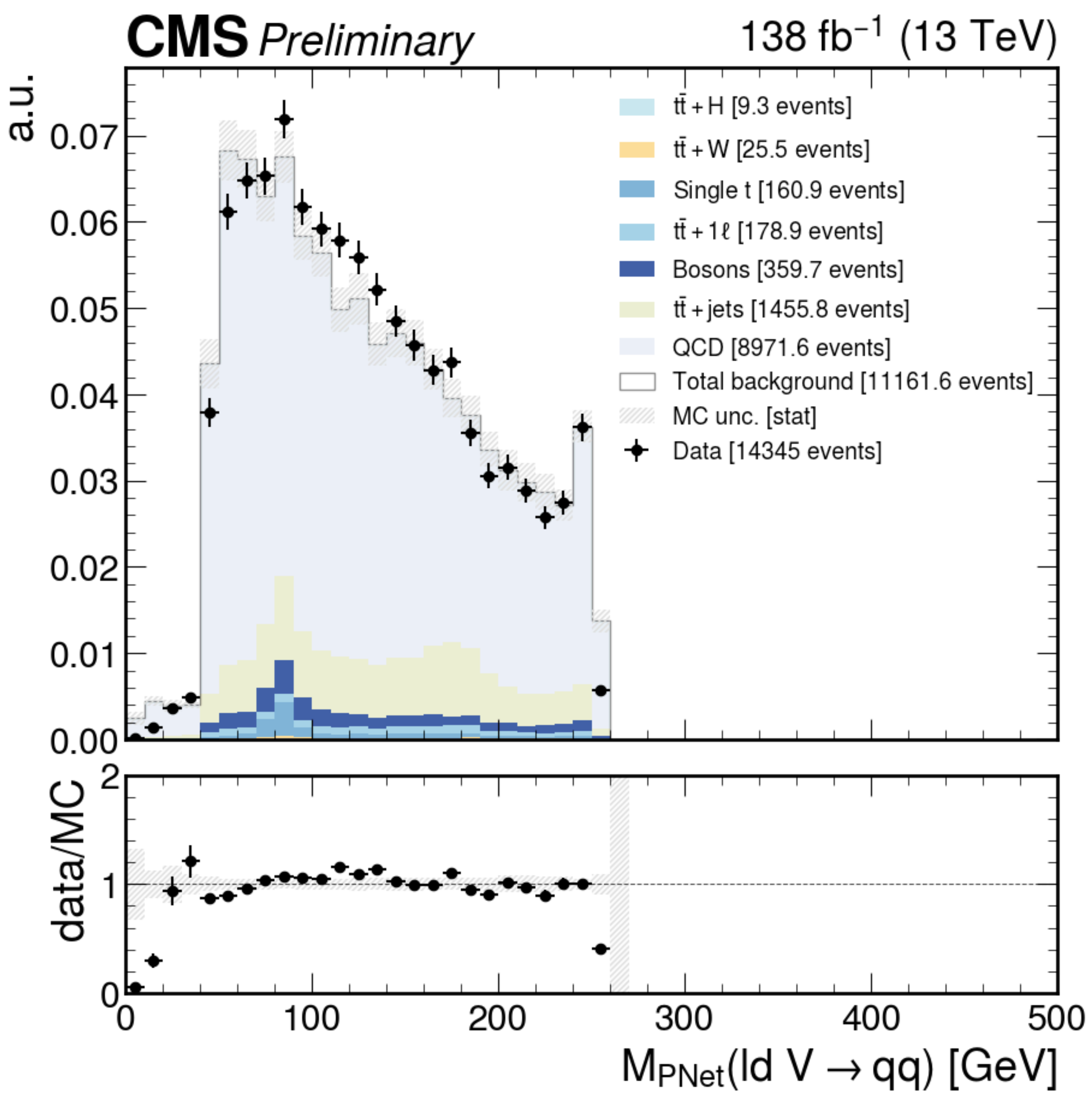


About 2x more weighted QCD events with correction after Preselection

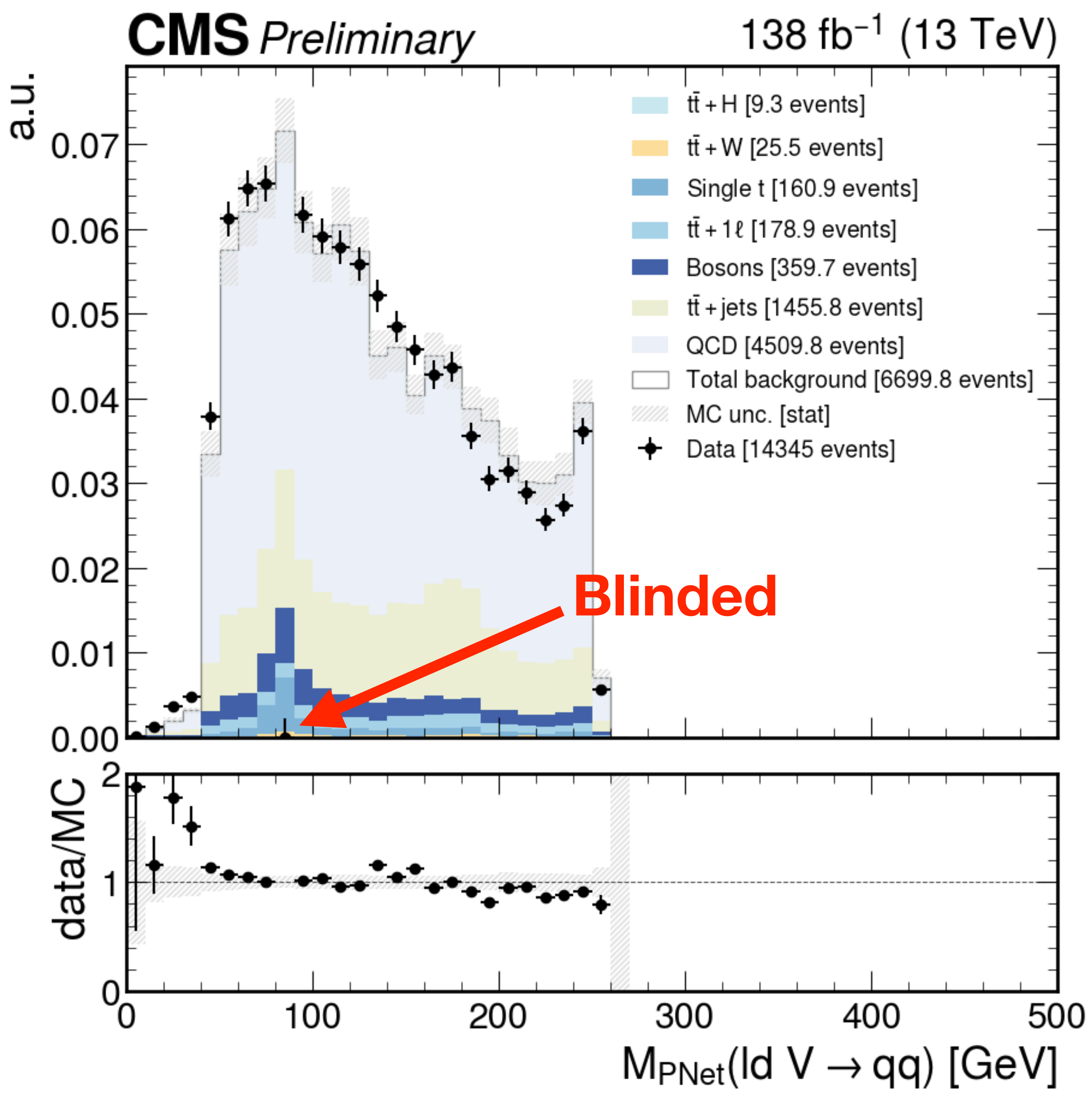
QCD integral is normalized to (data - non-QCD) after this step

QCD Corrections: With vs. Without

With



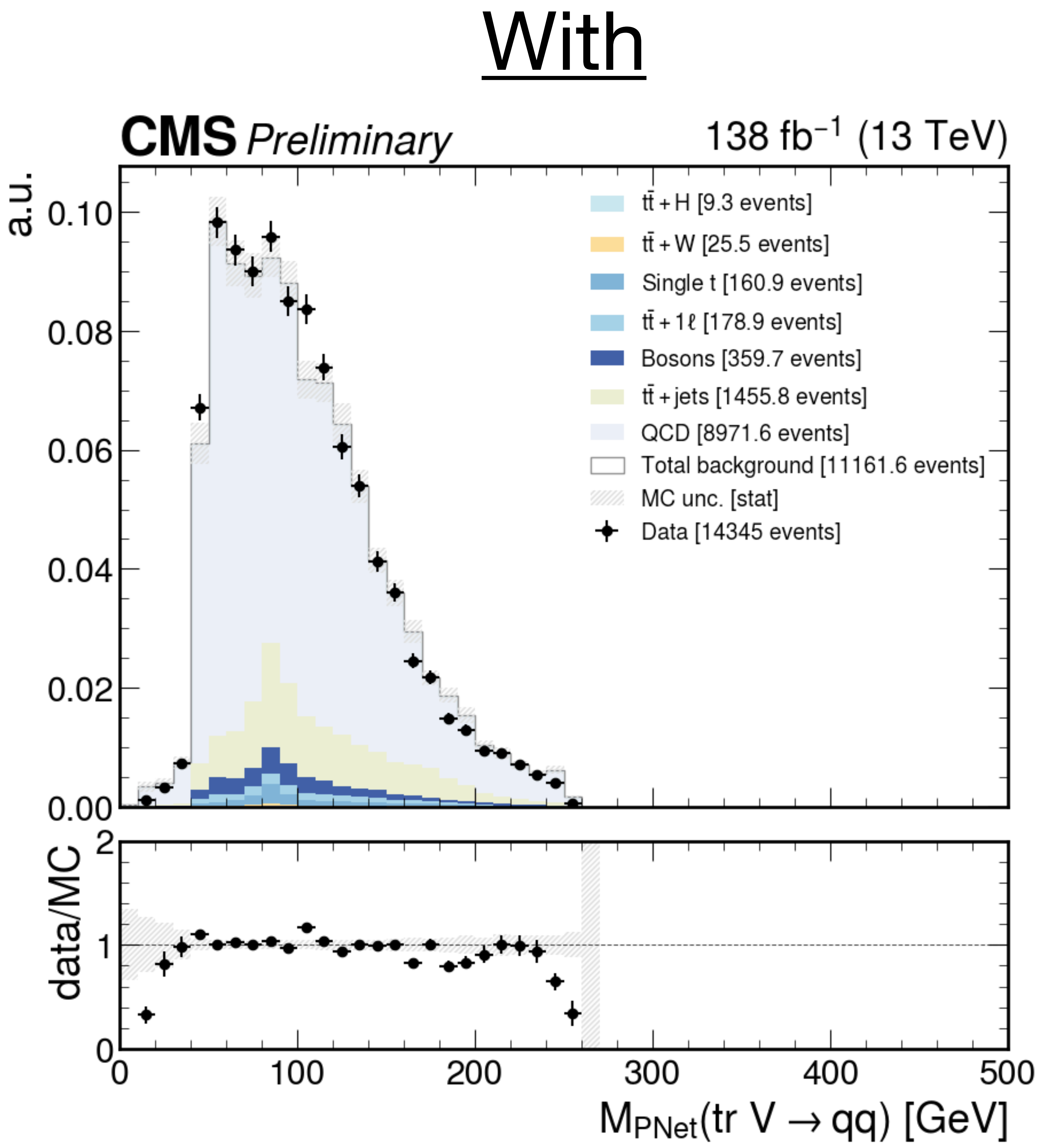
Without

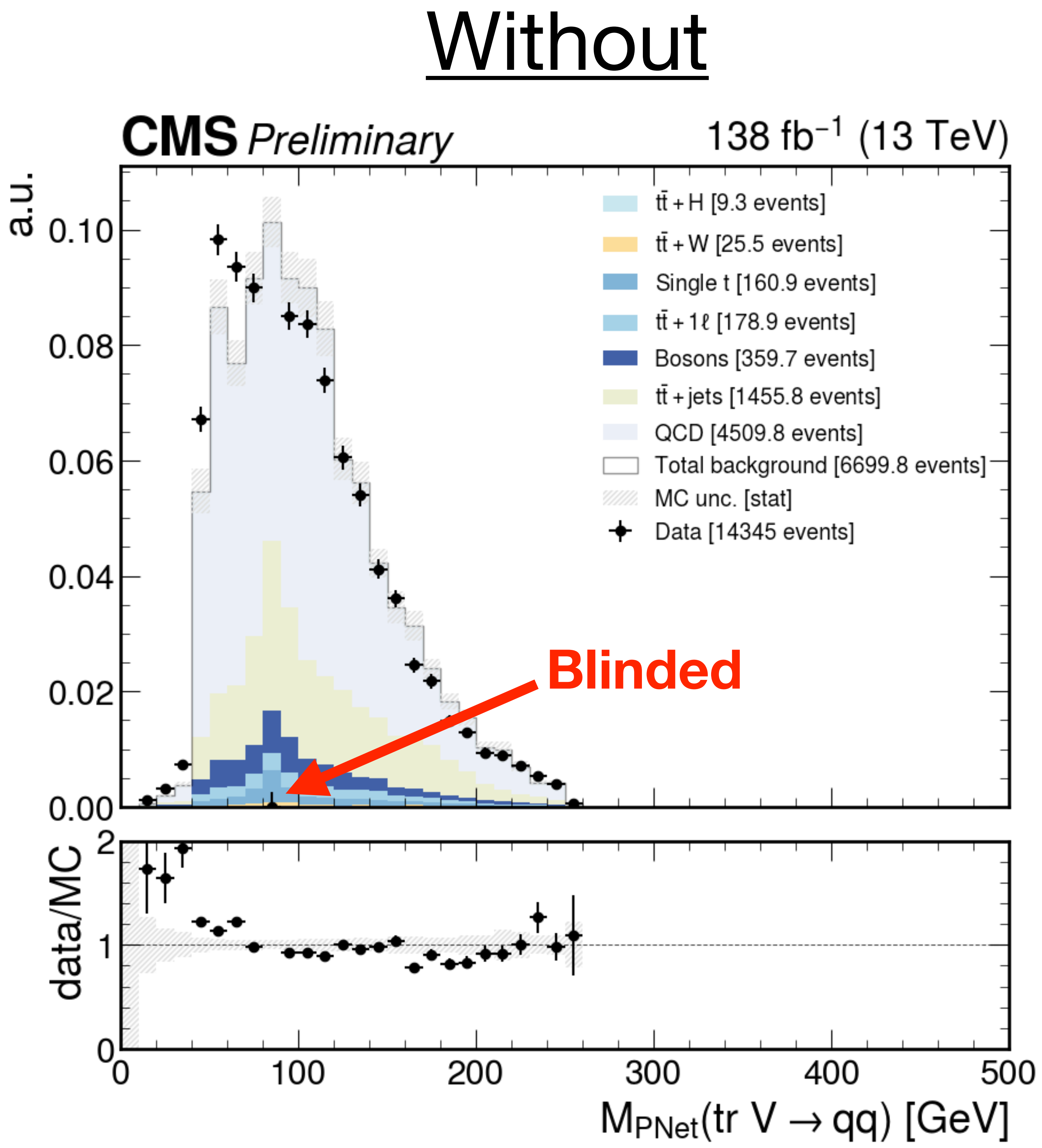


About 2x more weighted QCD events with correction after Preselection

QCD integral is normalized to (data - non-QCD) after this step

QCD Corrections: With vs. Without





About 2x more weighted QCD events with correction after Preselection

QCD integral is normalized to (data - non-QCD) after this step

UC San Diego

Sanity Check

- **Goal:** repeat the first example in the PRL paper (3D gaussian variables)
- (1) and (2) define the 3D gaussians
- (3) and (4) give the rest:
  - Input: X_1, X_2 (DisCo target: X_0)
  - NN architecture: 3 hidden layers; 128 nodes per layer; ReLU between layers; sigmoid output
  - λ = 1000, Adam optimizer
  - 2M sig, 2M bkg (batch size = 40K)

IV. APPLICATIONS

This section explores the efficacy of single and double DisCo in some applications of the ABCD method.

A. Simple example: Three-dimensional Gaussian random variables

We begin with a simple example to build some intuition and validate our methods. Consider a three-dimensional space (X_0, X_1, X_2), where the signal and background are both multivariate Gaussian distributions. We choose the means $\vec{\mu}$ and a covariance matrix $\mathbf{\Sigma}$ for background and signal as

1

$$\vec{\mu}_b = (0,0,0), \qquad \mathbf{\Sigma}_b = \sigma_b^2 \begin{pmatrix} 1 & \rho_b & 0 \\ \rho_b & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \sigma_b = 1.5, \qquad \rho_b = -0.8, \tag{4.1}$$

and

2

$$\vec{\mu}_s = (2.5, 2.5, 2), \qquad \mathbf{\Sigma}_s = \sigma_s^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \sigma_s = 1.5. \tag{4.2}$$

3

So for the background, all three features are centered at the origin and features X_0 and X_1 are correlated with each other but independent of X_2. For the signal, all three features are independent but are centered away from the origin. The first feature X_0 will play the role of the known feature for single DisCo in Sec. III.

All of the neural networks presented in this section use three hidden layers with 128 nodes per layer. The rectified linear unit (ReLU) activation function is used for the intermediate layers and the output is a sigmoid function. A hyperparameter of $\lambda = 1000$ is used for both single and double DisCo to ensure total decorrelation. The single DisCo training converged after 100 epochs while the double DisCo training required 200 epochs. Other networks only needed ten epochs. The double DisCo networks

were trained using a single neural network with a two-

4

dimensional output. All models were trained using Tensorflow [89] through Keras [90] with Adam [91] for optimization. Two million examples were generated with 15% used for testing. A batch size of 1% of the total was used for all networks to ensure an accurate calculation of the DisCo term in the relevant loss functions.

We first consider two classifiers: a baseline classifier $f_{\rm BL}(X_1, X_2)$ trained only on X_1 and X_2 and a single DisCo classifier $f_{\rm SD}(X_1, X_2)$ which includes a penalty for correlations between $f_{\rm SD}$ and X_0. The values of these classifiers for events drawn from the distributions are plotted in Fig. 3 against the X_0, X_1, or X_2 values of these events. We see that even though X_0 was not used in the training of the baseline, the classifier output is still correlated with X_0 because of the correlations between X_0 and X_1. In contrast to the baseline classifier, the single DisCo classifier is independent of both X_0 and X_1 and is simply a function of X_2. Intuitively, it makes sense that a classifier that must be independent of X_0 must also be independent of X_1. This is justified rigorously in Appendix B.

For double DisCo, we train two classifiers $f_{\rm DD}(X,Y,Z)$ and $g_{\rm DD}(X,Y,Z)$ according to the double DisCo loss function. The results are illustrated in Fig. 4. The first classifier depends mostly on Z and the second classifier depends mostly on X and Y. However, the residual dependence on all three observables is not a deficit of the training procedure: even though the three random variables are separable into two independent subsets (X,Y) and Z, the two classifiers learned by double DisCo

CMS

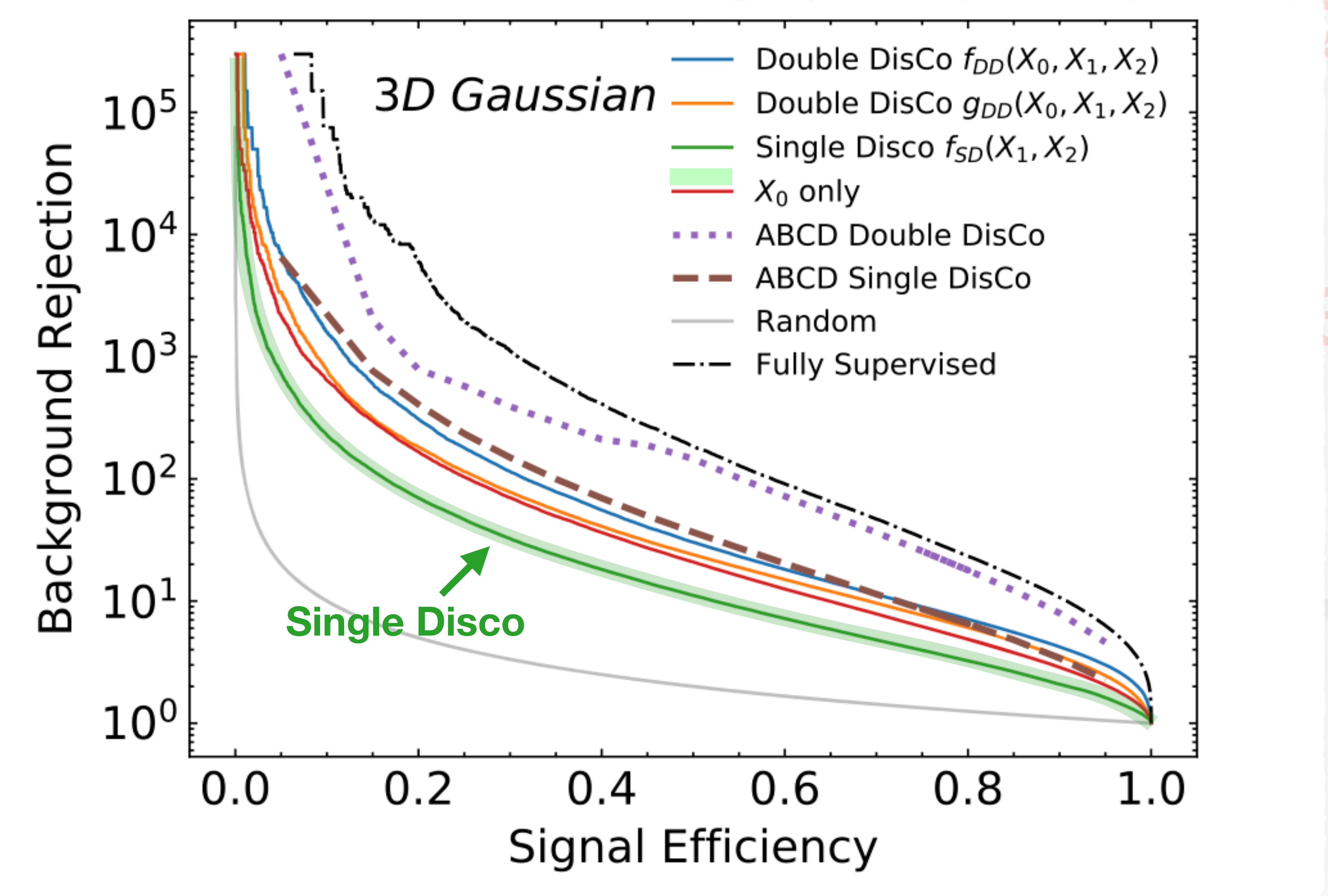
UC San Diego

Sanity Check

- **Goal:** repeat the first example in the PRL paper (3D gaussian variables)
- (1) and (2) define the 3D gaussians
- (3) and (4) give the rest:
  - Input: X_1, X_2 (DisCo target: X_0)
  - NN architecture: 3 hidden layers; 128 nodes per layer; ReLU between layers; sigmoid output
  - λ = 1000, Adam optimizer
  - 2M sig, 2M bkg (batch size = 40K)

Target: recreate their plots, e.g.

1/FPR

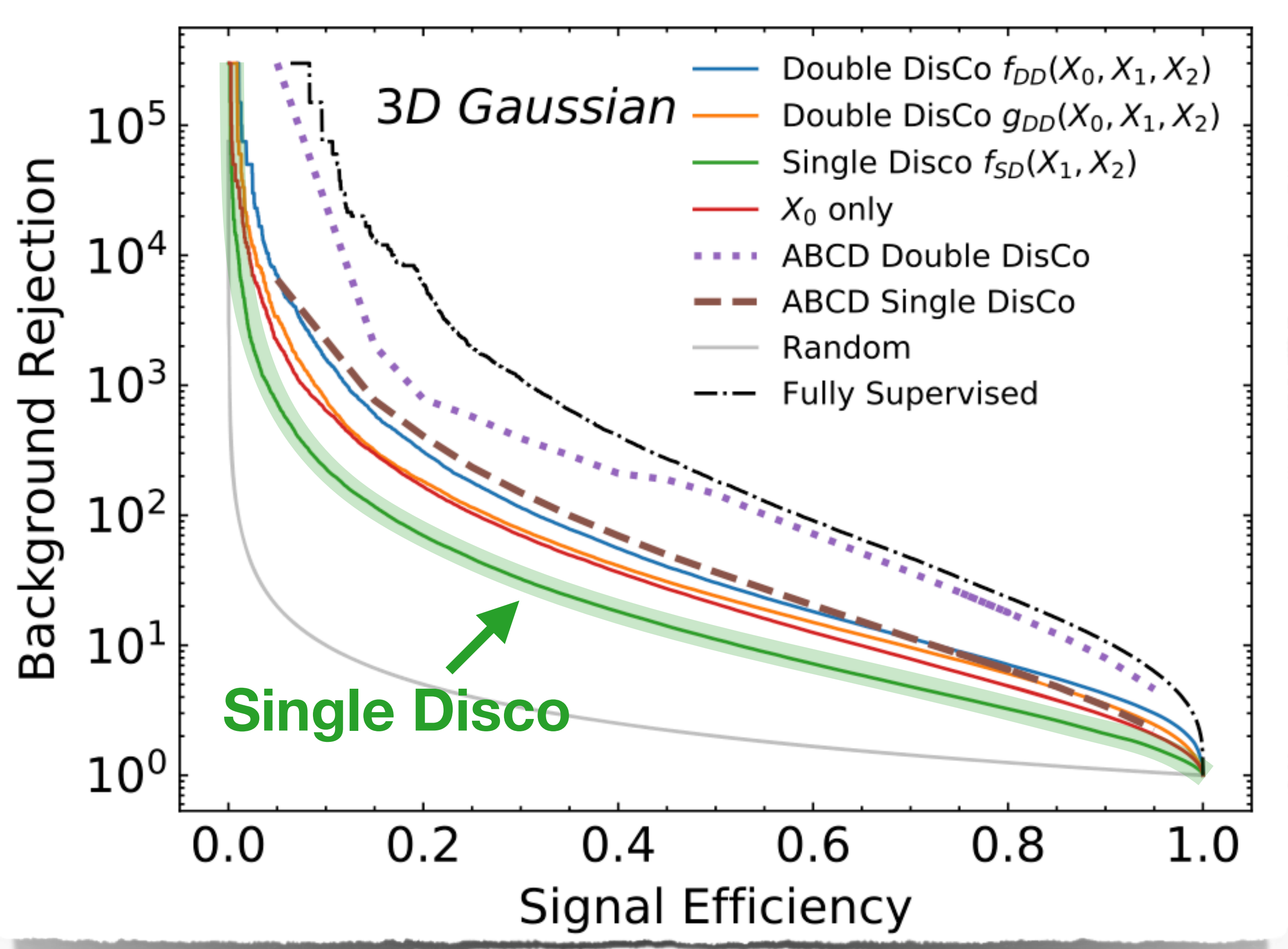


TPR

TPR = TP/P = (true positives)/(positives)
FPR = FP/N = (false positives)/(negatives)

3D Gaussians: λ = 1000 DisCo

$$\mathcal{L} = \mathcal{L}_{BCE}(f_{SD}(X_1, X_2), y) + 1000 \times \text{dCorr}_{y=0}(f_{SD}(X_1, X_2), X_0)$$



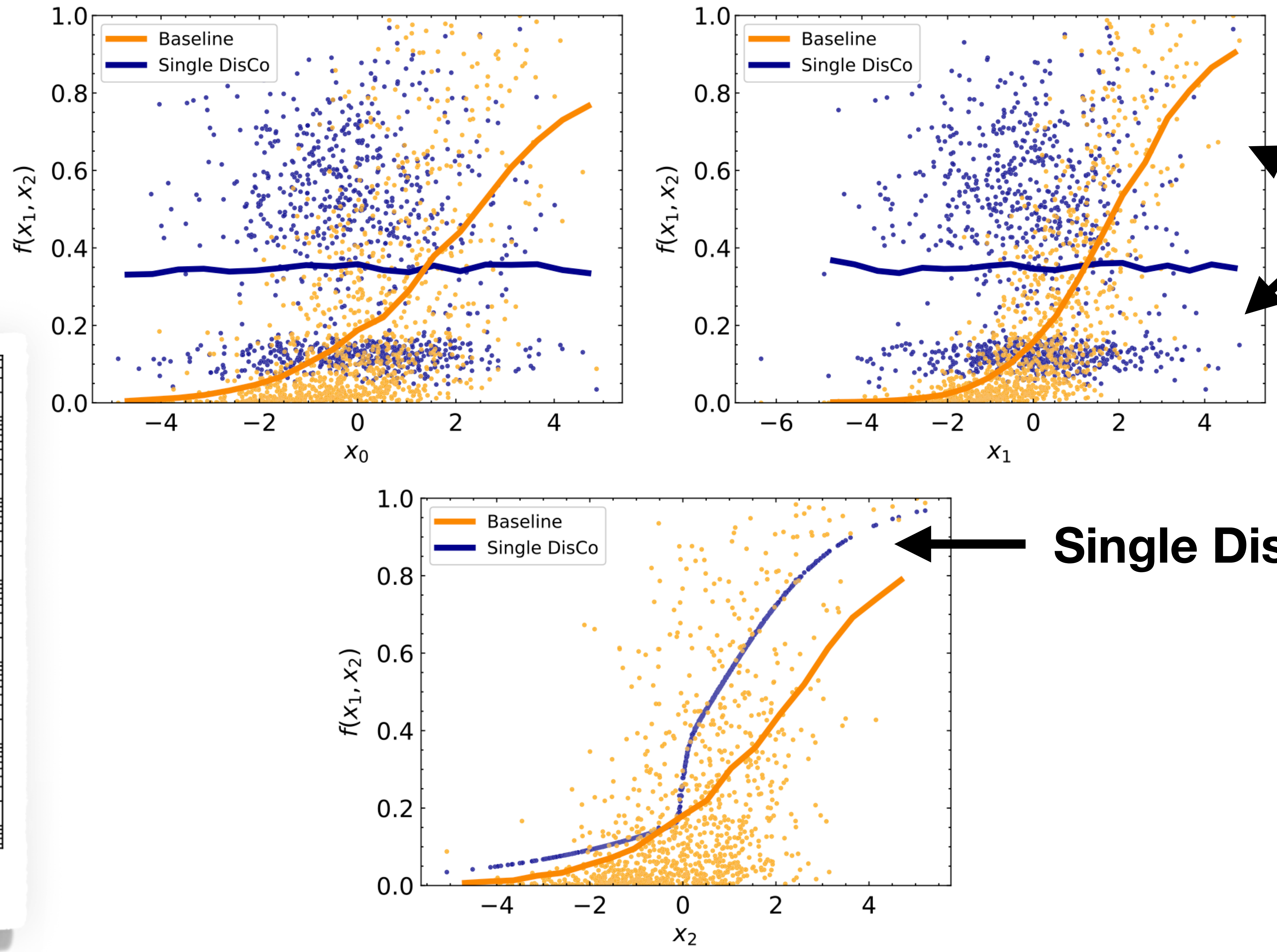


FIG. 3. Scatter plots showing the relationship (or lack thereof) between the three random variables X_0, X_1, and X_2 and (1) a baseline classifier $f_{\mathrm{BL}}(X_1, X_2)$ trained on X_1 and X_2 with no regularization, and (2) a classifier $f_{\mathrm{SD}}(X_1, X_2)$ trained with the single DisCo loss function that penalizes correlations with X_0. Only the background events are shown in these plots. The solid lines are the averages of the classifiers over events with the same value of X_0, X_1, or X_2. In the third panel, the scatter of the single DisCo classifier is already a line, so no average is needed.

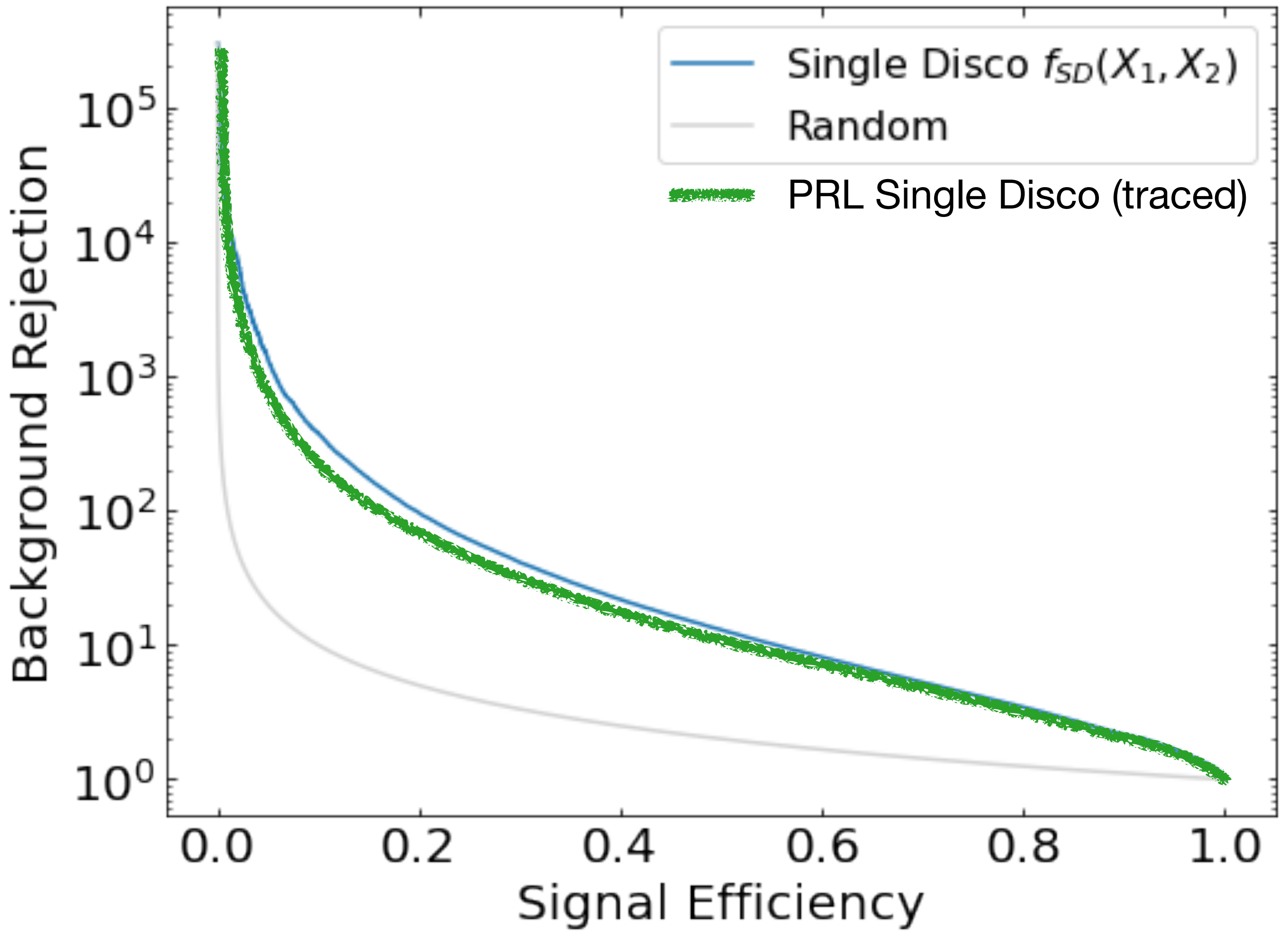
Two distinct clusters

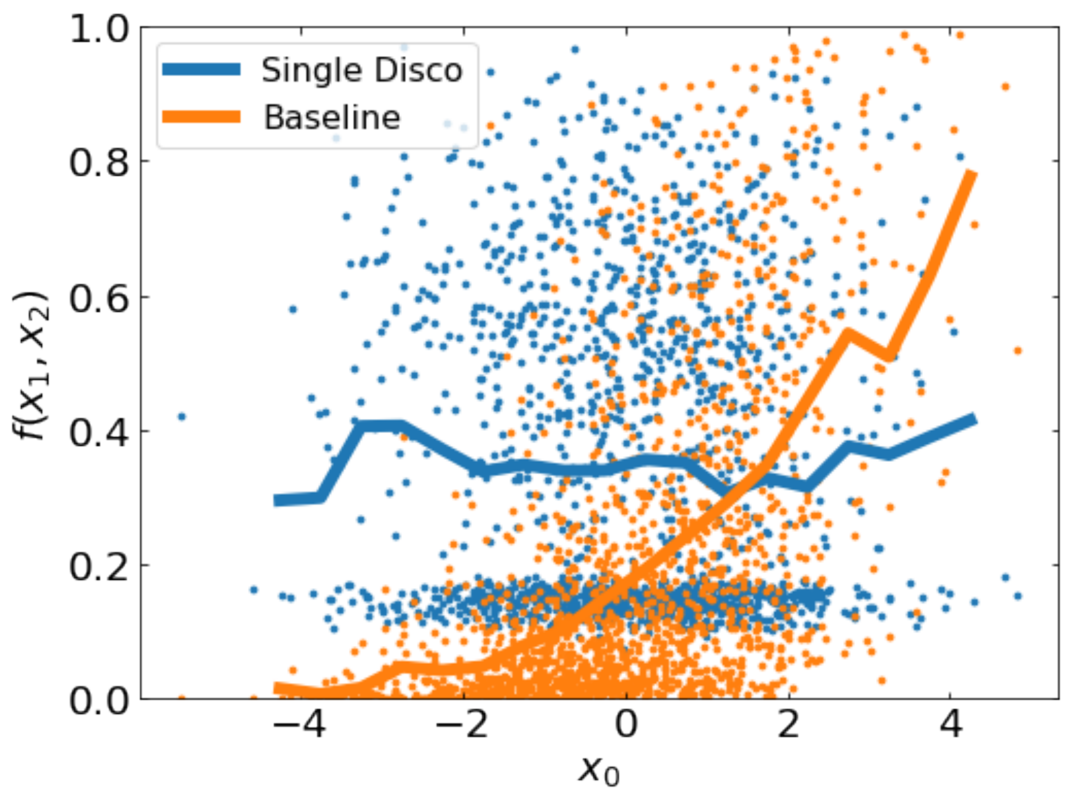
Single DisCo is a perfect line

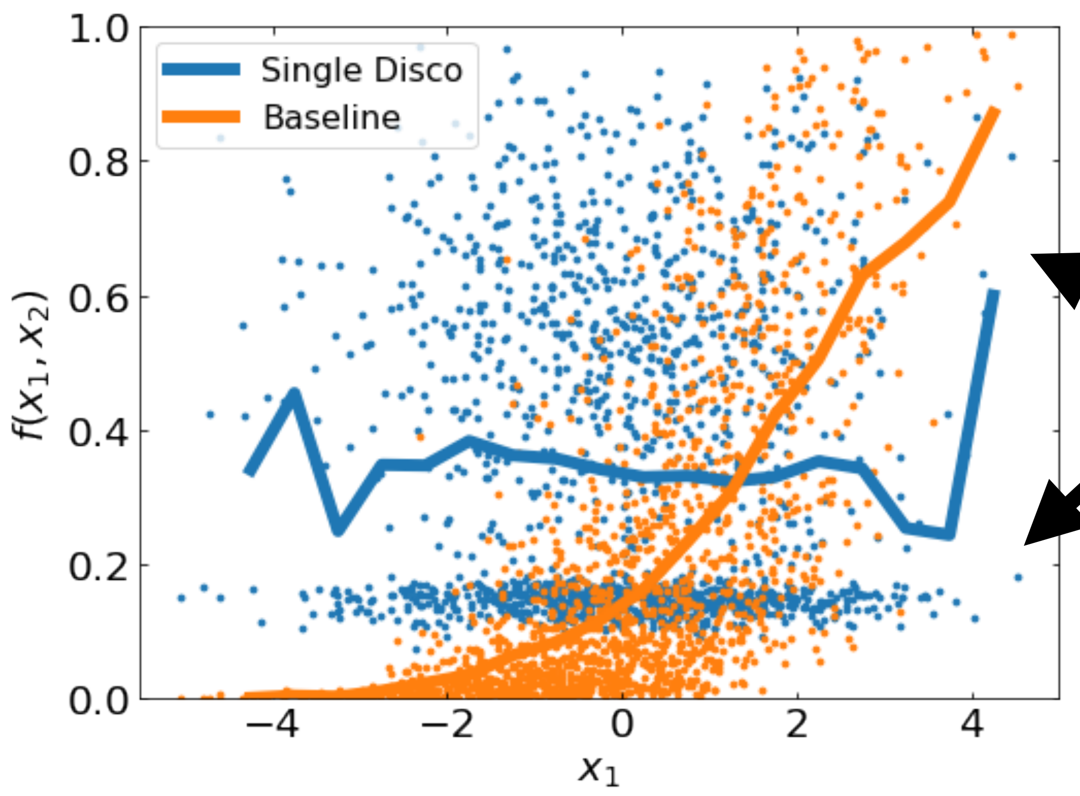
3D Gaussians: λ = 1000 DisCo

$$\mathscr{L} = \mathscr{L}_{BCE}(f_{SD}(X_1, X_2), y) + 1000 \times \mathrm{dCorr}_{y=0}(f_{SD}(X_1, X_2), X_0)$$

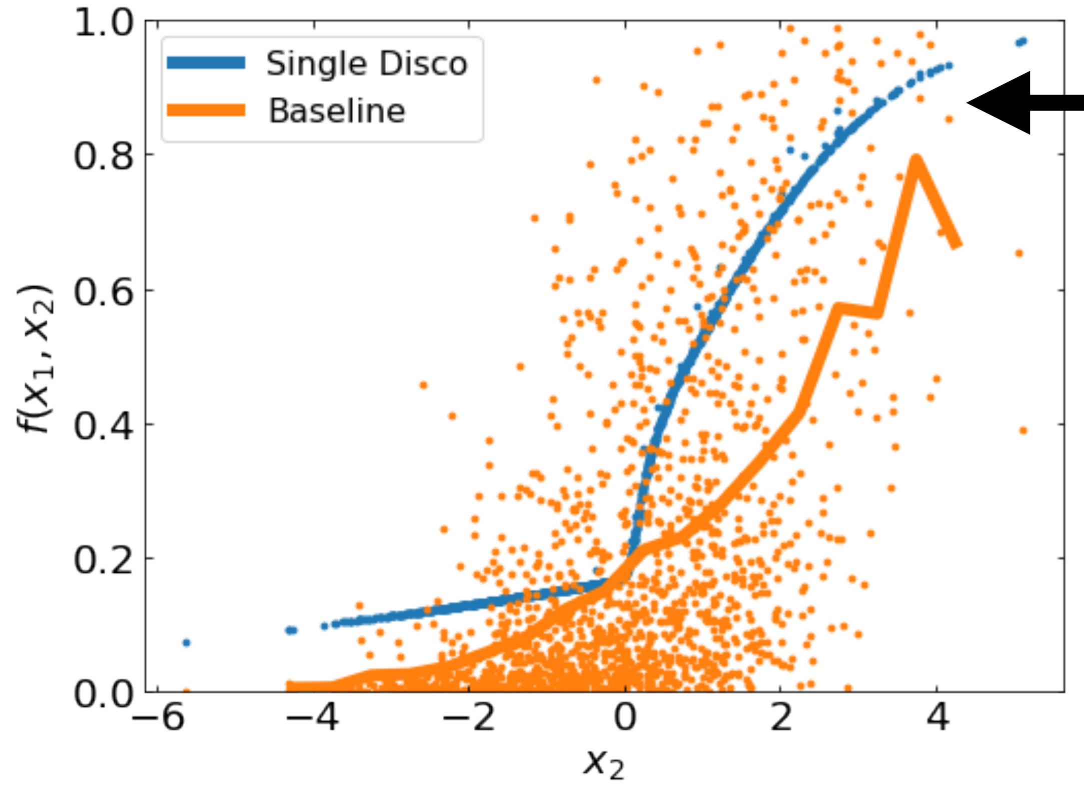
My ROC curve matches ✓







Two distinct clusters ✓



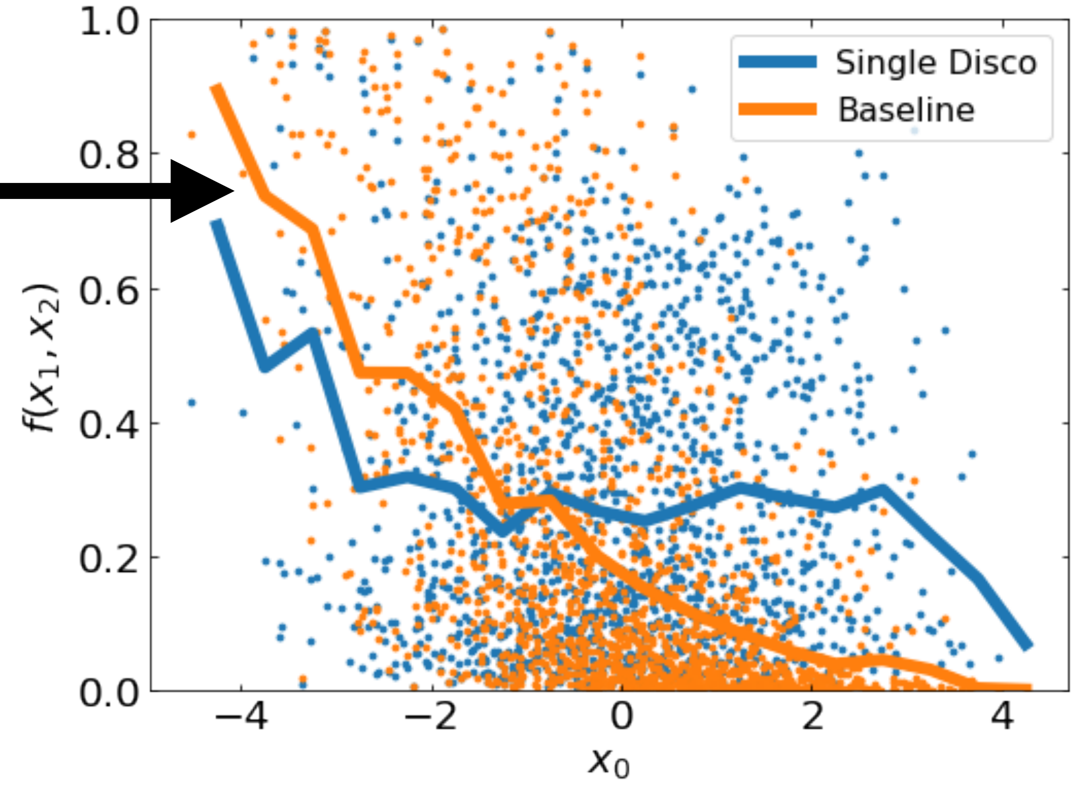
Single DisCo is a perfect line ✓

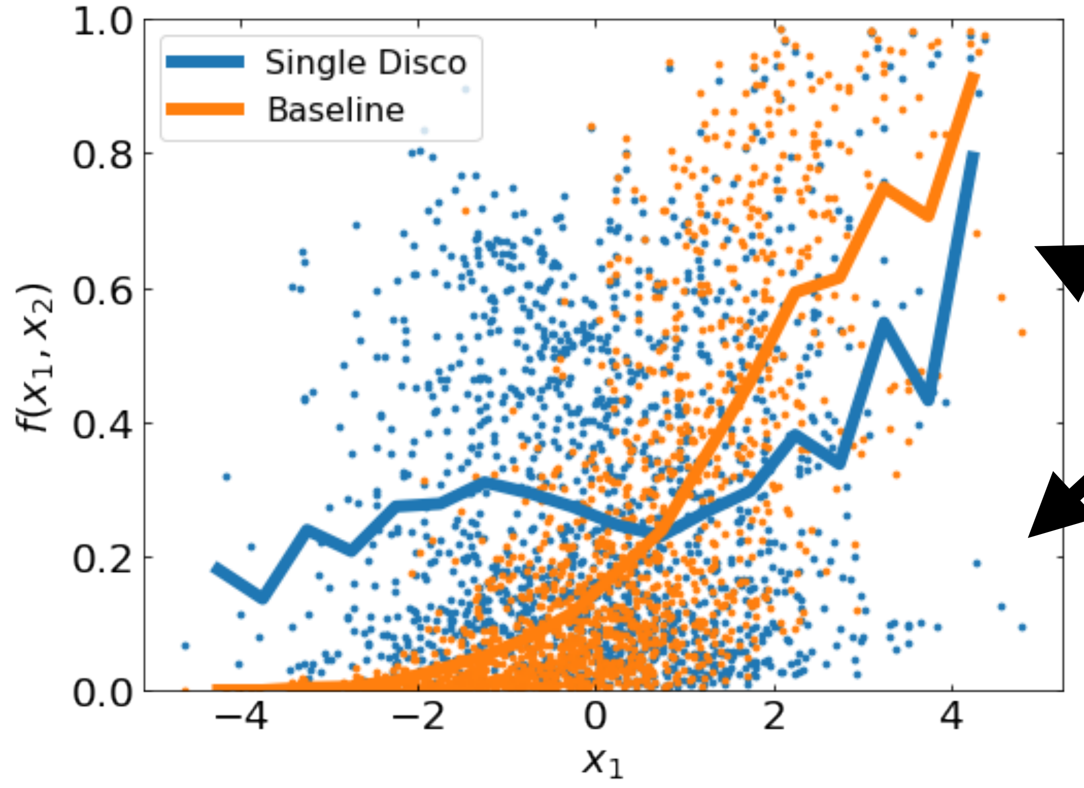
My plots match those in the PRL paper ✓

3D Gaussians: λ = 1000 DisCo

Originally, we were unable to reproduce 3D Gaussian example

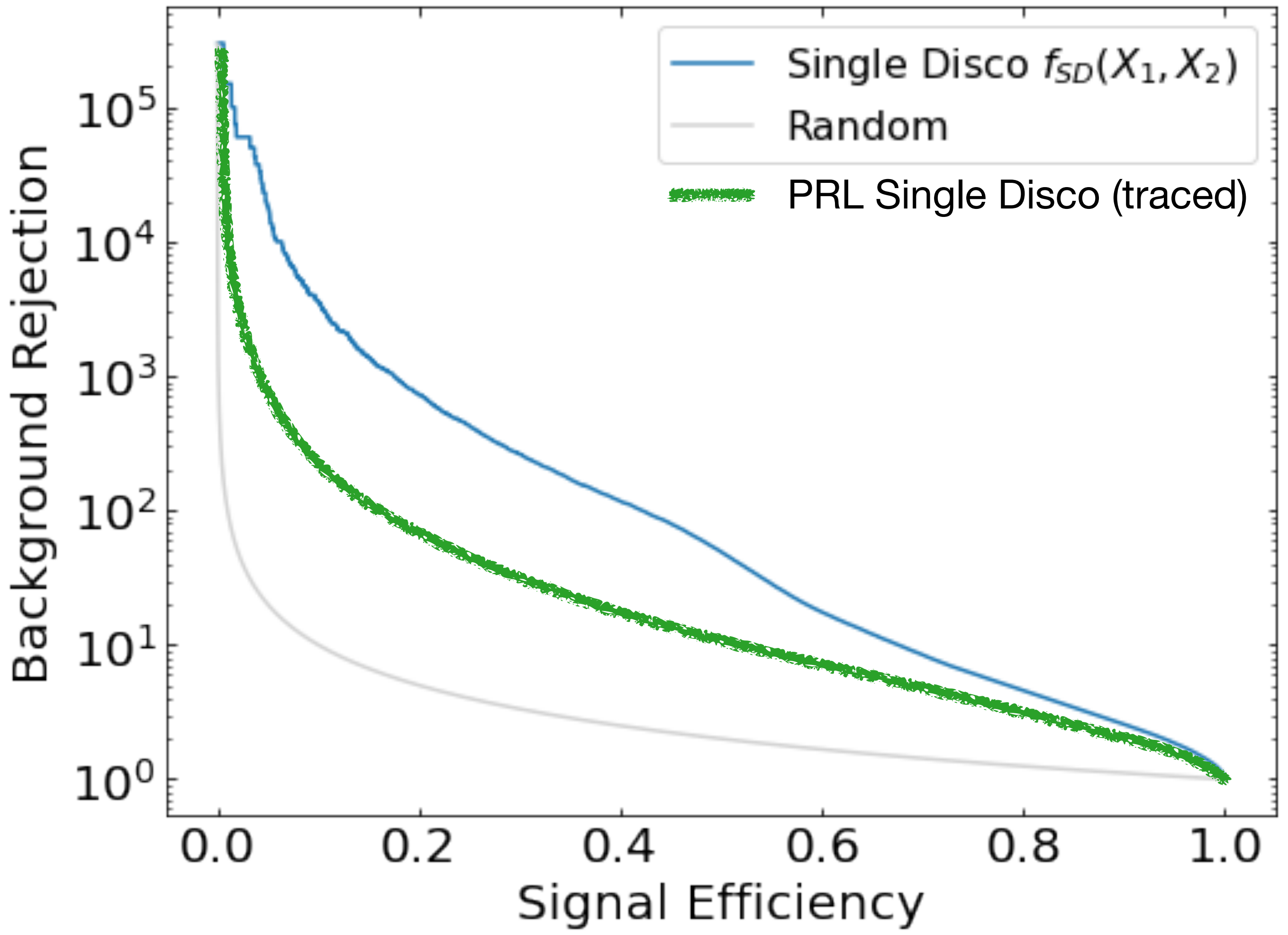
Baseline avg. opposite of that in the PRL paper

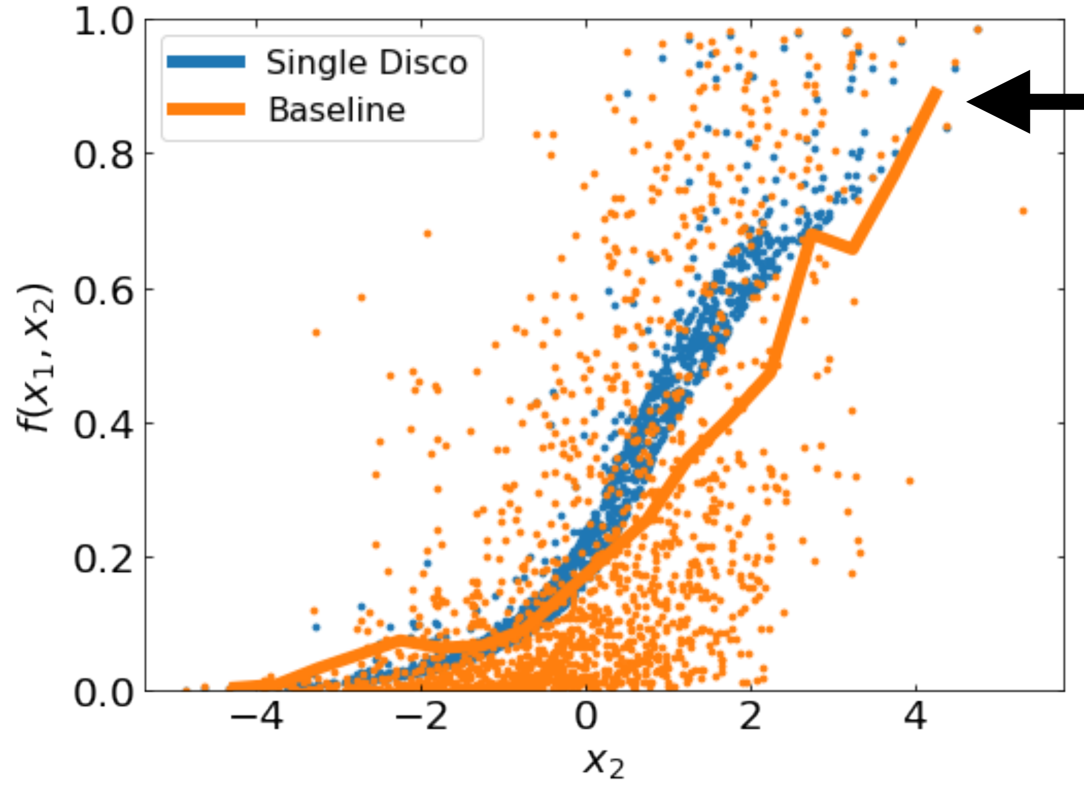




No distinct clusters

My ROC curve is different



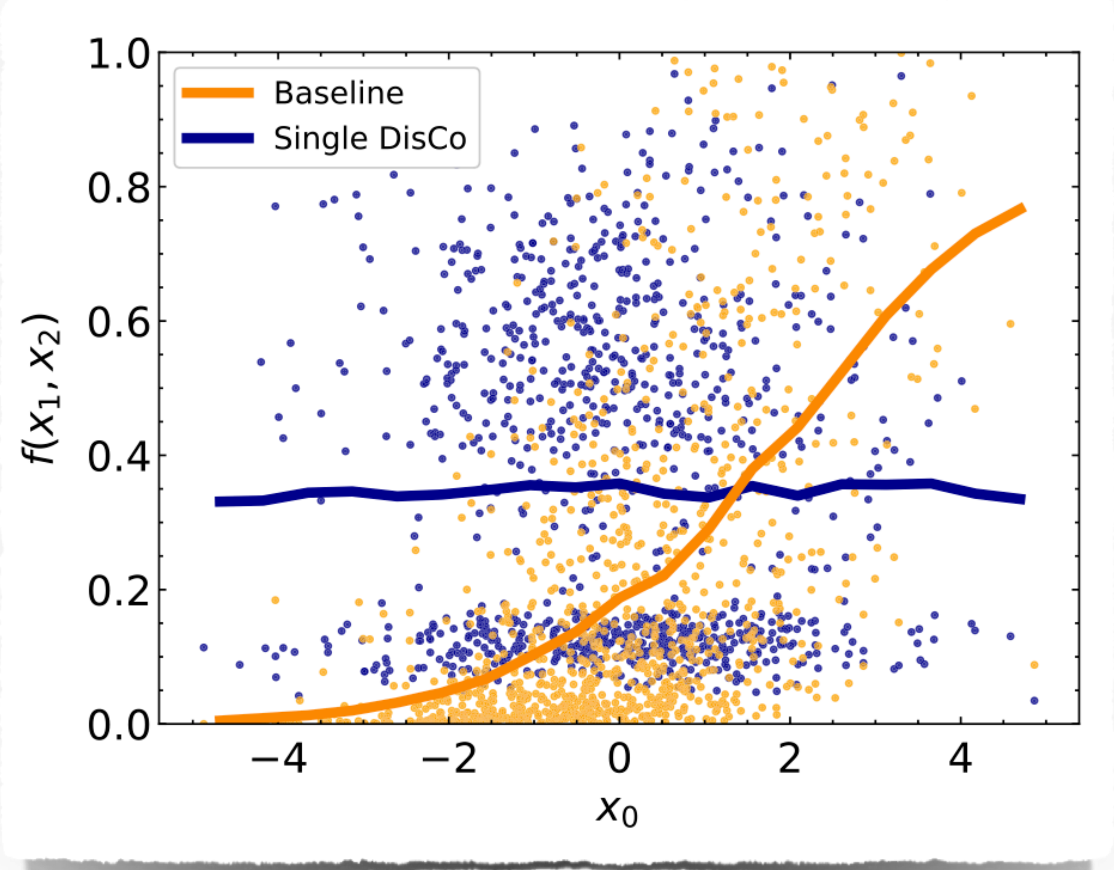


Single DisCo not a perfect line (more similar to Baseline avg.)

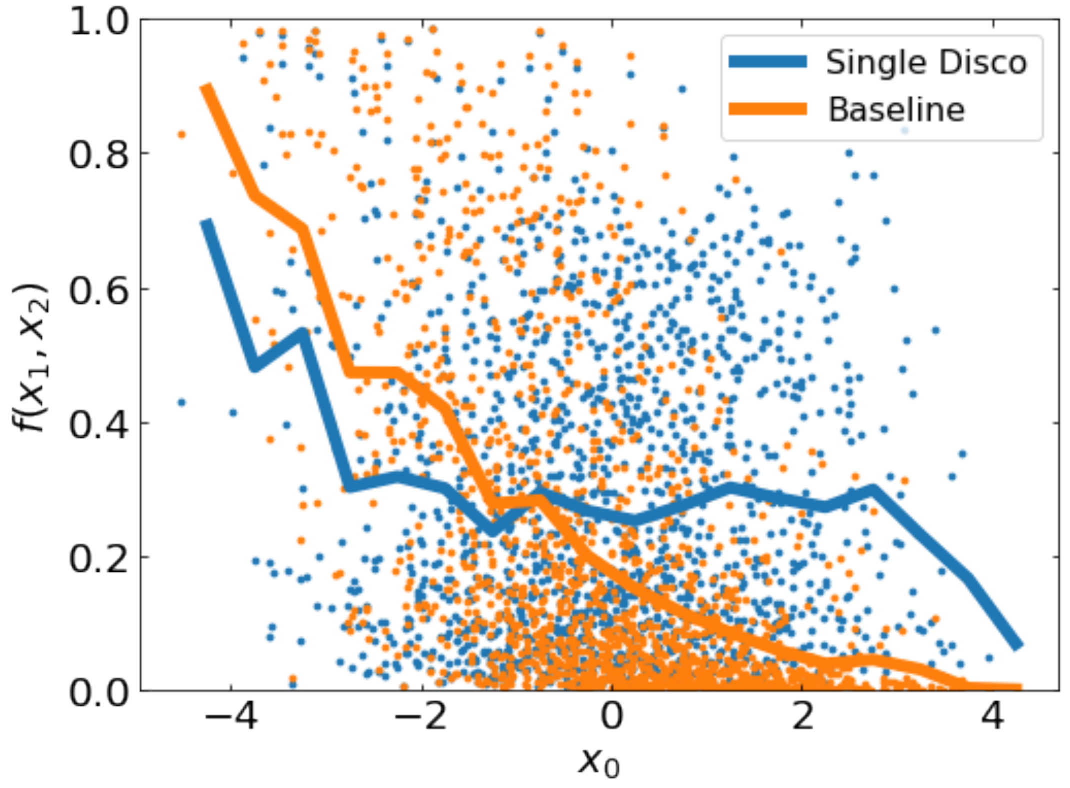
3D Gaussians: λ = 1000 DisCo

Spotted two issues (typos?) with PRL paper

Fig. 3



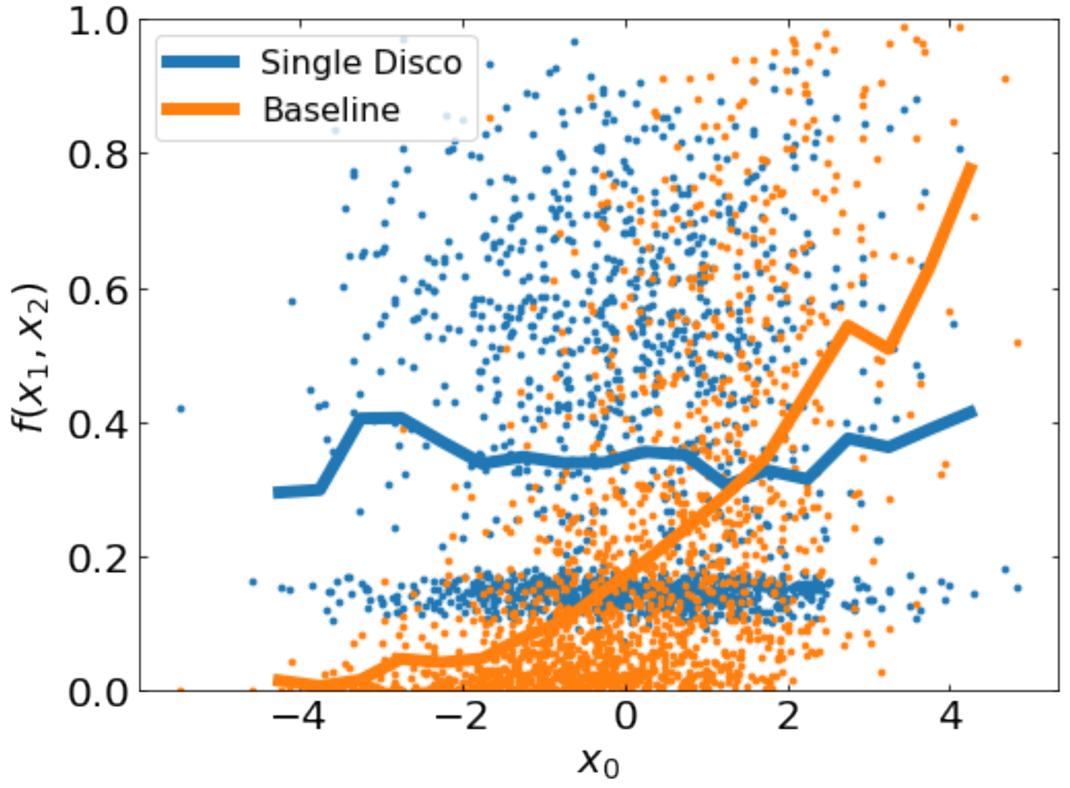
vs.



Does not match PRL

dCorr² → dCorr

Set ρ = +0.8



Matches PRL ✓

Eq. 3.1

$$\mathcal{L}[f(X)] = \mathcal{L}_{\text{classifier}}[f(X), y] + \lambda \text{dCorr}^2_{y=0}[f(X), X_0],$$

dCorr² → dCorr

$$\lambda \text{dCorr}_{y=0}[f(X), X_0]$$

Eq. 4.1

$$\vec{\mu}_b = (0,0,0), \qquad \Sigma_b = \sigma_b^2 \begin{pmatrix} 1 & \rho_b & 0 \\ \rho_b & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -0.8 & 0 \\ -0.8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\sigma_b = 1.5, \qquad \rho_b = -0.8,$$

Set ρ = +0.8

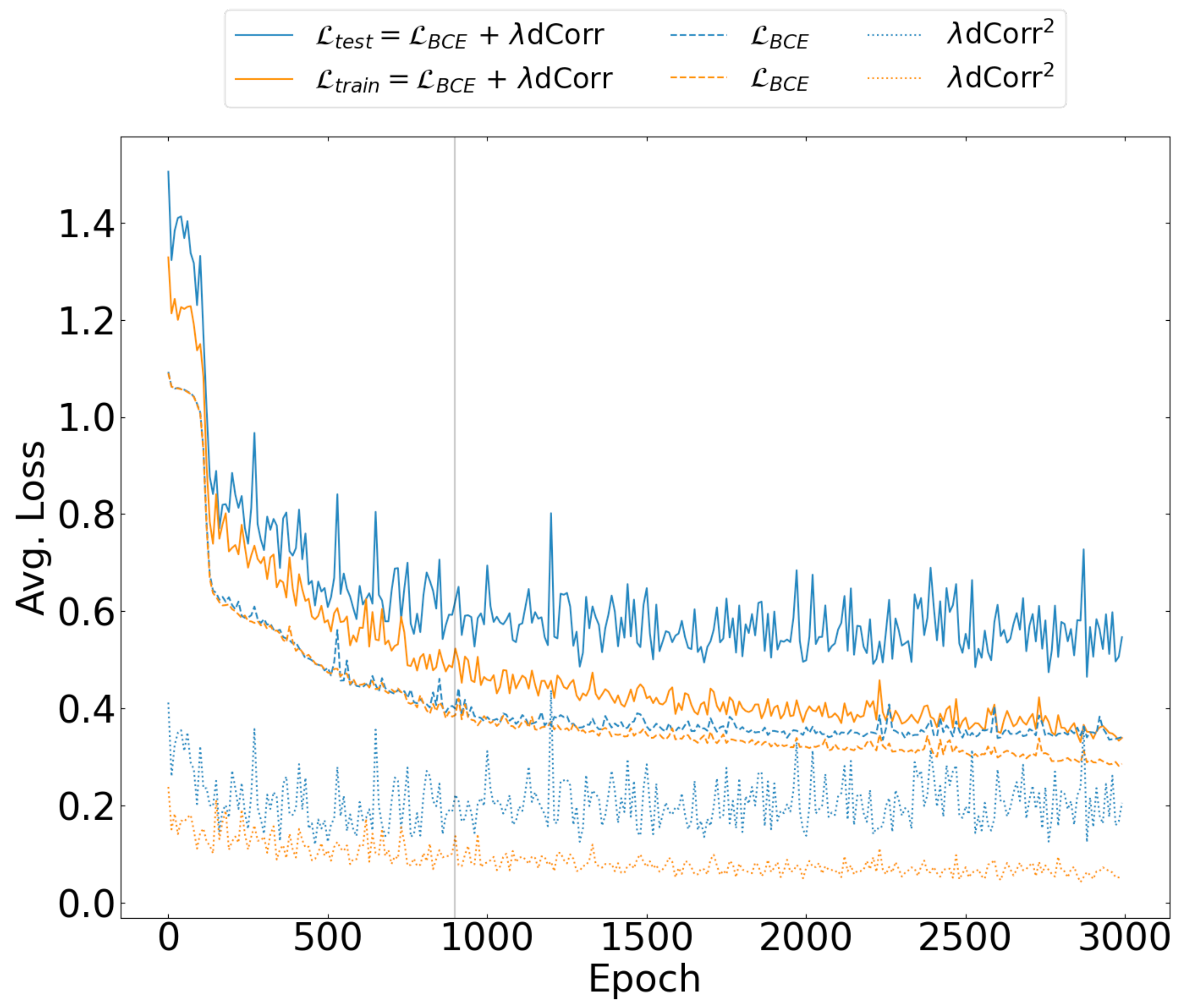
$$\Sigma_b = \sigma_b^2 \begin{pmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

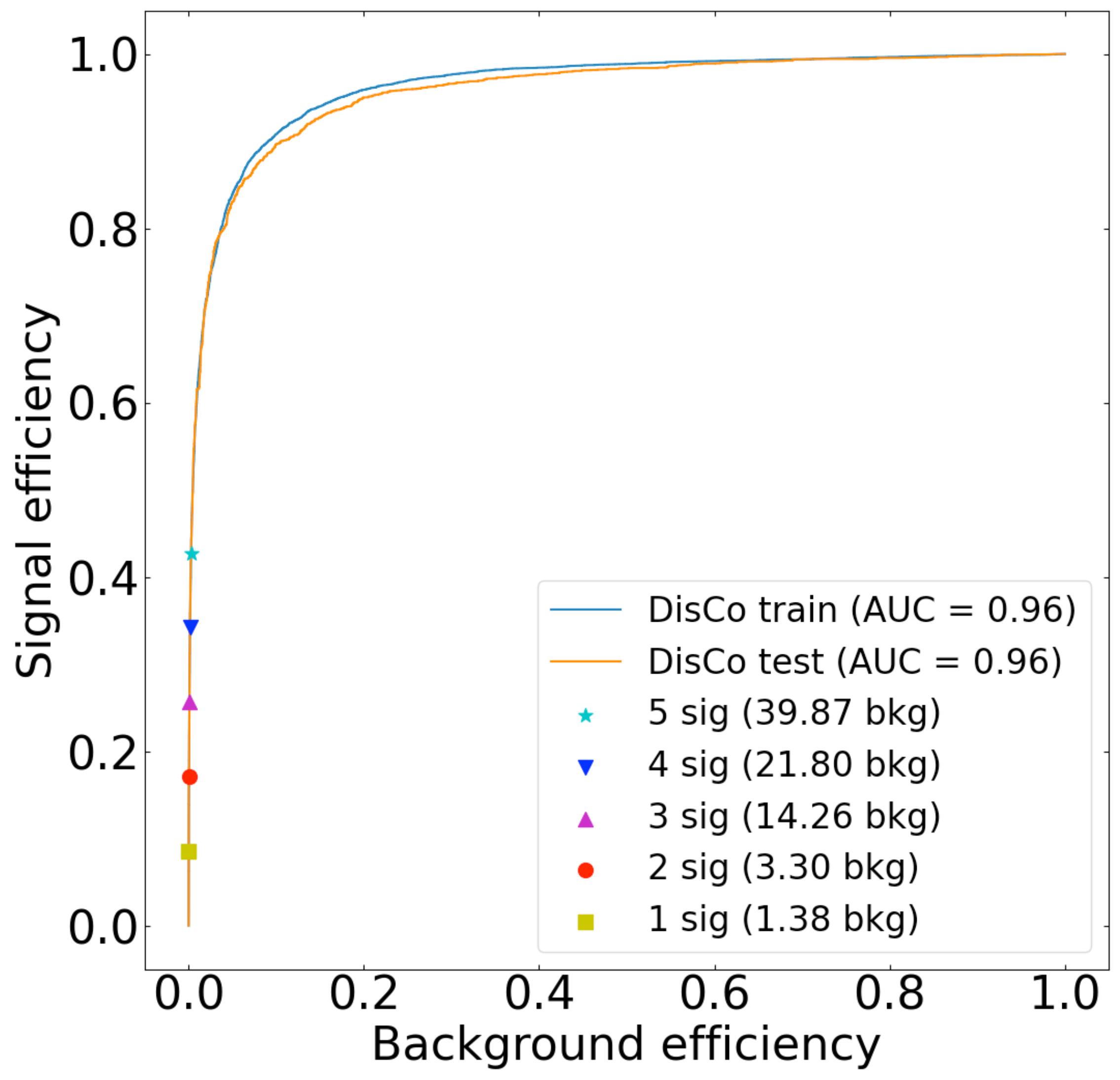
UC San Diego

ABCDNet: Training Details

$$\mathscr{L} = \mathscr{L}_{BCE}(f(\vec{x}), y) + 30 \times \text{dCorr}_{y=0}(f(\vec{x}), |\Delta\eta_{jj}|)$$

- 3 hidden layers (64 nodes each)
  - Leaky ReLU for hidden layers
  - Sigmoid for output layer
- Input features:
  - $H \to bb$ fat jet p4 (i.e. p_T, η, ϕ), M_{PNet}
  - $V \to qq$ fat jet p4 (i.e. p_T, η, ϕ), M_{PNet}
  - M_{jj}
- $p_T \to \ln(p_T)$, others $\to$ (x - min)/(max - min)



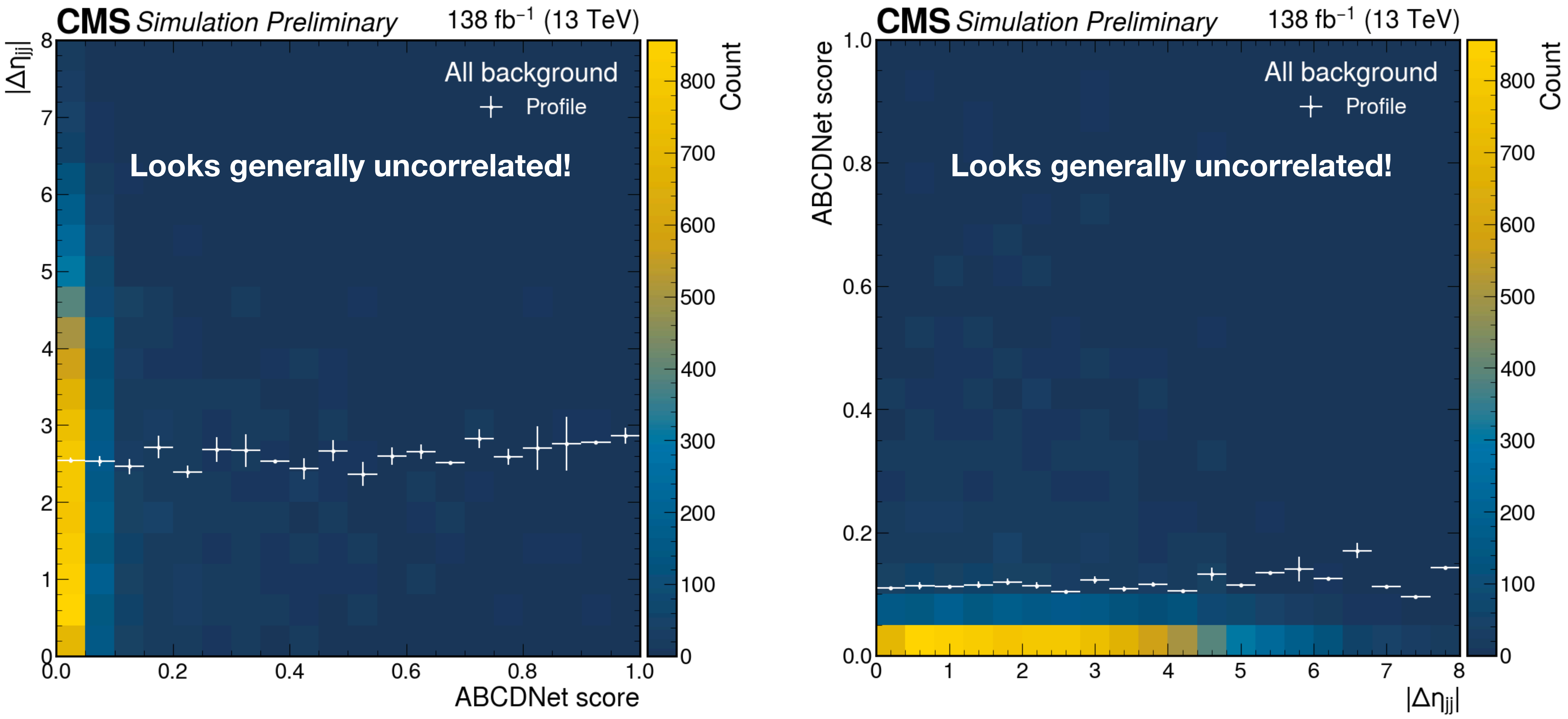


Tried a number of λ values:

λ > 30 destabilizes training (fails to converge)

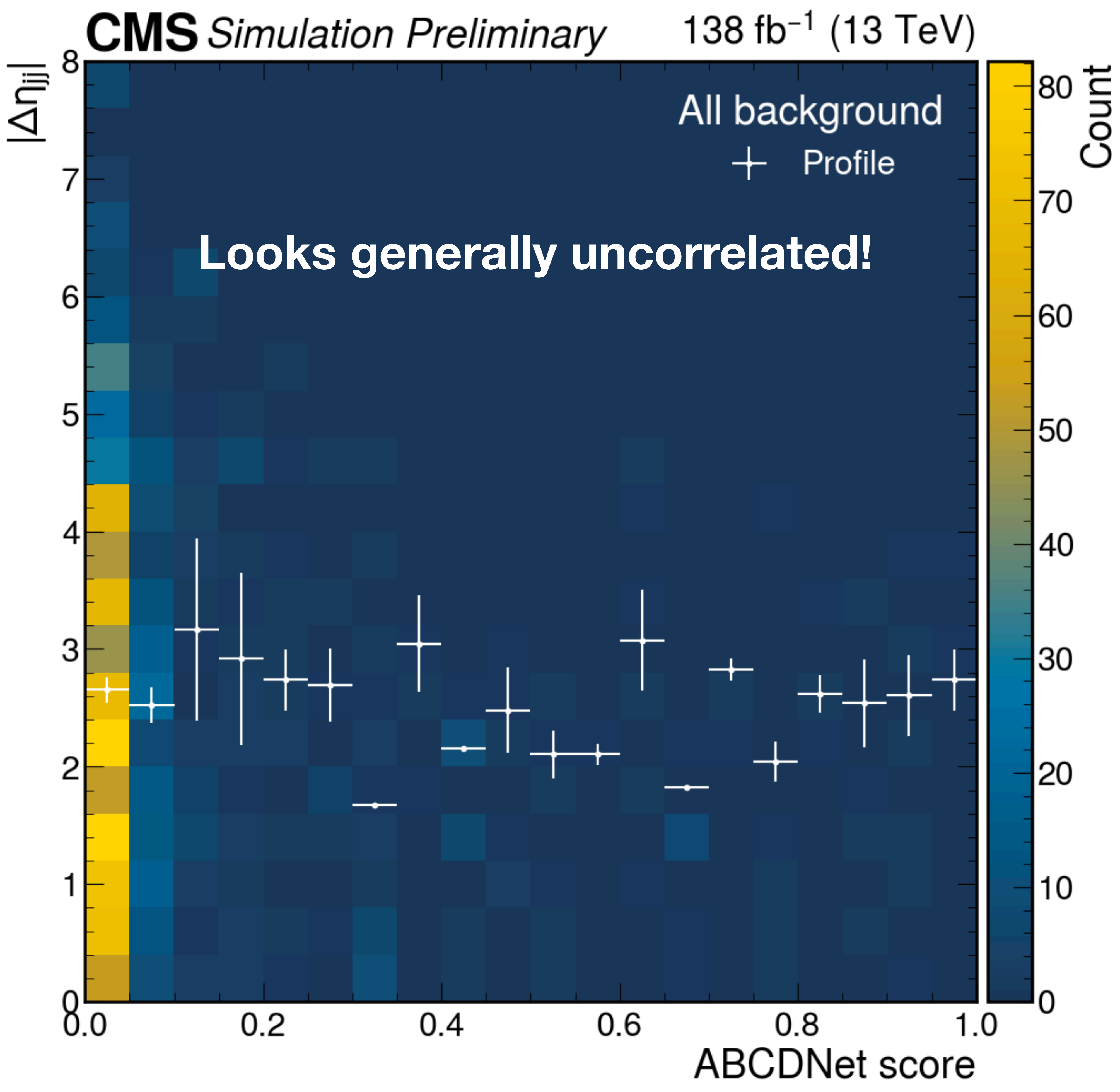
λ < 30 misses decorrelation (poor ABCD closure)

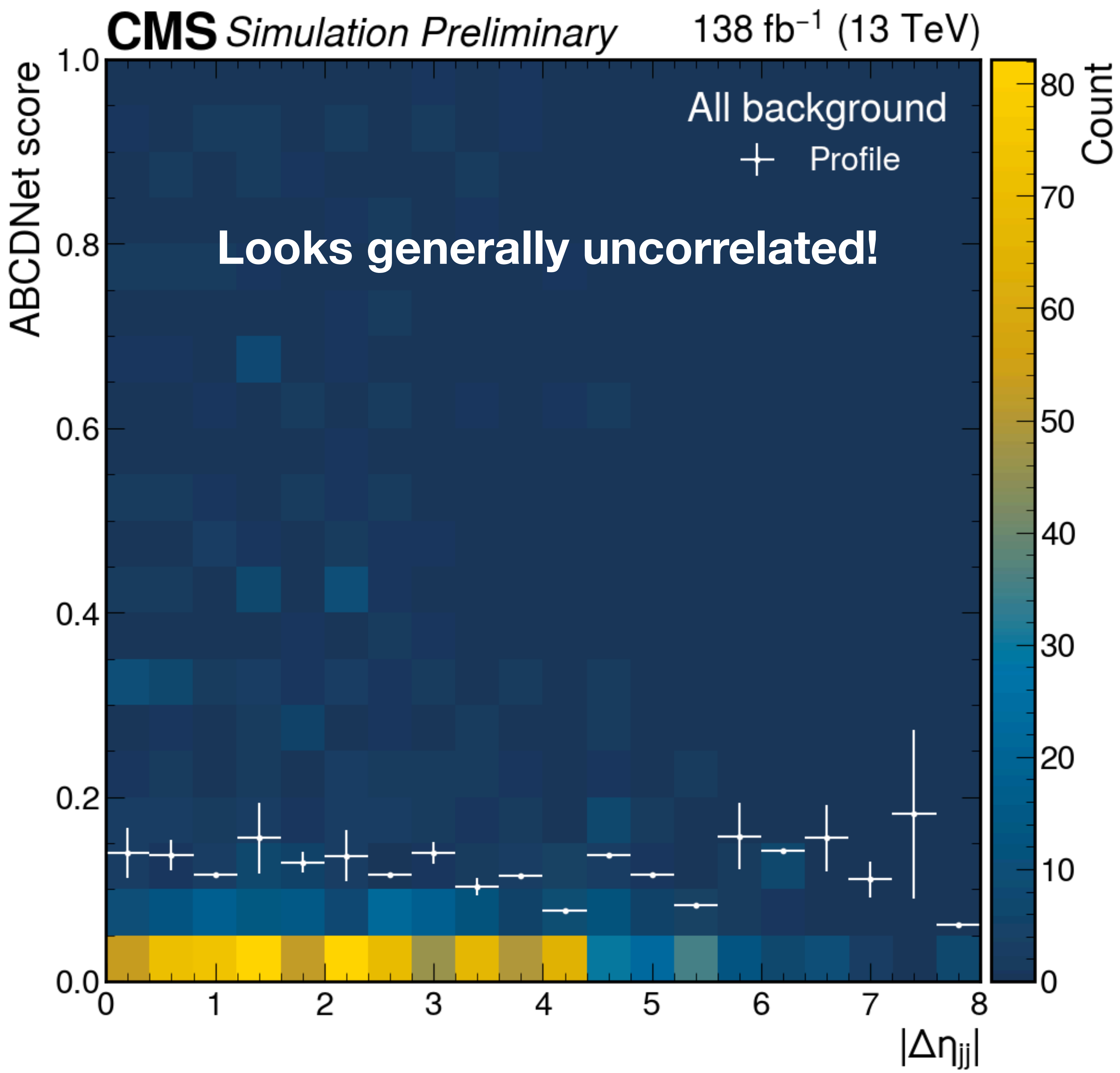
ABCDNet: Decorrelation Results



Selection: Preselection

ABCDNet: Decorrelation Results





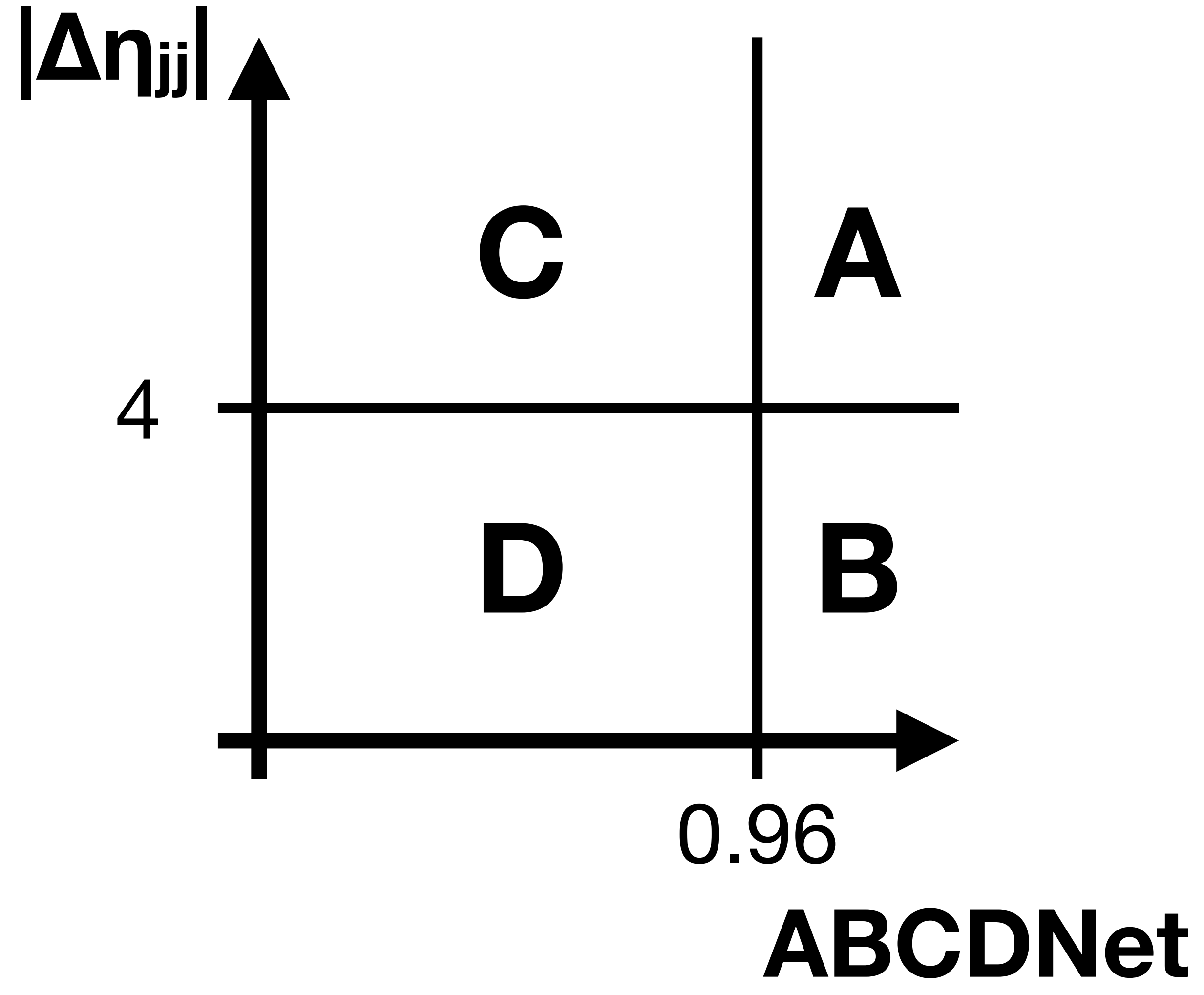
Selection: Xbb > 0.6 and XWqq > 0.75|0.70 (ld|tr)

ABCDNet: ABCD Bkg. Composition

Xbb > 0.60 and XWqq > 0.75|0.70 (ld|tr)

Region	Bosons	err.	%	QCD	err.	%	Single top	err.	%	$t\bar{t}+1\ell$	err.	%	$t\bar{t}$+H	err.	%	$t\bar{t}$+jets	err.	%	$t\bar{t}$+W	err.	%	Total	err.
A	**0.43**	0.16	30%	**0.67**	0.48	46%	**0.15**	0.15	10%	**0.14**	0.08	9%	**0.00**	0.00	0%	**0.07**	0.07	5%	**0.00**	0.00	0%	1.46	0.57
B	**0.68**	0.23	13%	**3.33**	0.98	65%	**0.57**	0.29	11%	**0.00**	0.00	0%	**0.01**	0.00	0%	**0.52**	0.20	10%	**0.01**	0.00	0%	5.12	1.25
C	**10.1**	1.07	3%	**214.0**	25.8	73%	**13.1**	1.37	4%	**5.63**	0.52	2%	**0.42**	0.03	0%	**48.4**	1.84	17%	**1.30**	0.35	0%	292.9	25.9
D	**44.0**	2.01	4%	**730.7**	40.8	72%	**32.8**	2.17	3%	**21.8**	1.02	2%	**1.63**	0.05	0%	**176.9**	3.54	17%	**4.51**	0.76	0%	1012.4	41.1

C/D	Bosons	QCD	Single top	$t\bar{t}+1\ell$	$t\bar{t}$+H	$t\bar{t}$+jets	$t\bar{t}$+W	Total
C/D	**0.23** ± 0.03	**0.29** ± 0.04	**0.40** ± 0.05	**0.26** ± 0.03	**0.26** ±0.02	**0.27** ±0.01	**0.29** ± 0.09	**0.29** ± 0.03



- **SR matches Preselection composition better than BDT SR**
- Many of the percentages above have large error bars:

Region	Bosons	QCD	Single top	$t\bar{t}+1\ell$	$t\bar{t}$+H	$t\bar{t}$+jets	$t\bar{t}$+W
A	30% ± 56%	46% ± **91%**	10% ± 22%	9% ± 18%	0% ± 1%	5% ± 10%	0% ± 0%
B	13% ± 13%	65% ± 62%	11% ± 12%	0% ± 0%	0% ± 0%	10% ± 10%	0% ± 0%
C	3% ± 1%	73% ± 27%	4% ± 2%	2% ± 1%	0% ± 0%	17% ± 6%	0% ± 0%
D	4% ± **1%**	72% ± 15%	3% ± 1%	2% ± 0%	0% ± 0%	17% ± 4%	0% ± 0%

$$\epsilon_{\%QCD} = (\%\text{QCD}) \times \sqrt{2\times\left(\frac{\epsilon_{QCD}}{A_{QCD}}\right)^2 + \left(\frac{\epsilon_{Bosons}}{A_{Bosons}}\right)^2 + \ldots + \left(\frac{\epsilon_{t\bar{t}+W}}{A_{t\bar{t}+W}}\right)^2}$$

$$\epsilon_{\%Bosons} = (\%\text{Bosons}) \times \sqrt{\left(\frac{\epsilon_{QCD}}{D_{QCD}}\right)^2 + 2\times\left(\frac{\epsilon_{Bosons}}{D_{Bosons}}\right)^2 + \ldots + \left(\frac{\epsilon_{t\bar{t}+W}}{D_{t\bar{t}+W}}\right)^2}$$

Epoch = 900 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego