Comparing Analysis Strategies

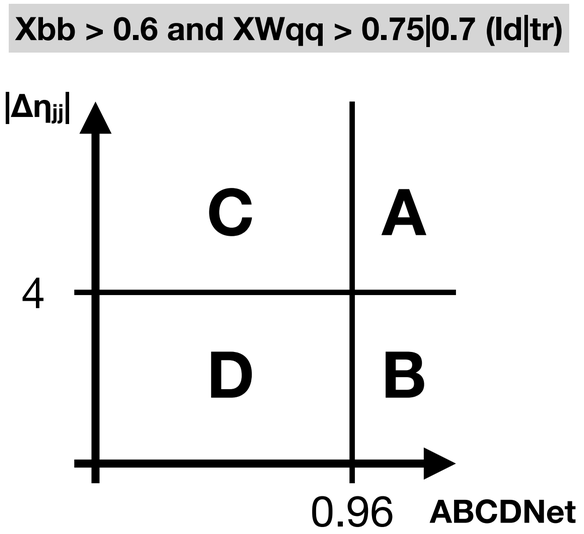
A. Arora, R. Band, P. Chang, L. Giannini, **J. Guiang**, C. Jessop, L. Lutton,
F. Würthwein, Y. Xiang, A. Yagil, Y. Zhou, L. Zygala

August 14th, 2023

UC San Diego
UF UNIVERSITY of FLORIDA
UNIVERSITY OF NOTRE DAME

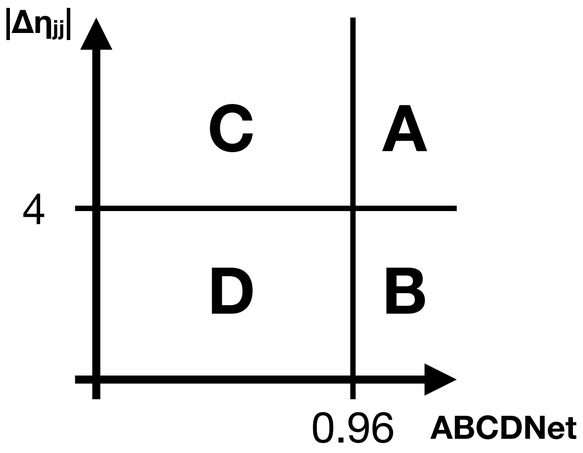
Overview

- ABCD shown to L3s (right) had limited statistics in Region B
- This resulted in a systematic uncertainty of ~50%
- Trying two more strategies:
  - ABCDNet vs. VBS BDT
  - ABCDNet vs. VBS DNN
- Then we compare: **is there a clear best strategy?**

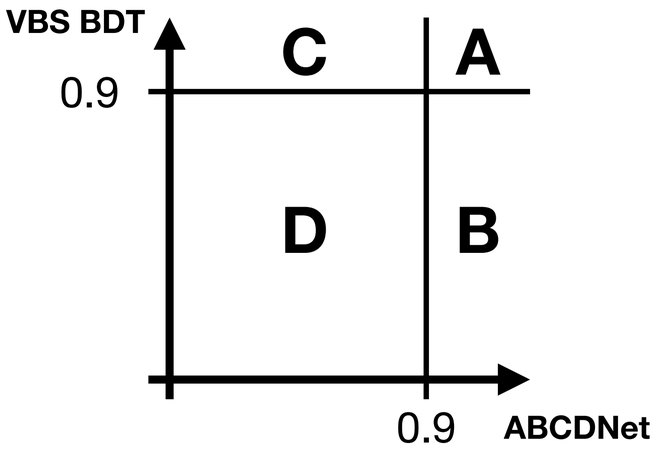


Comparison

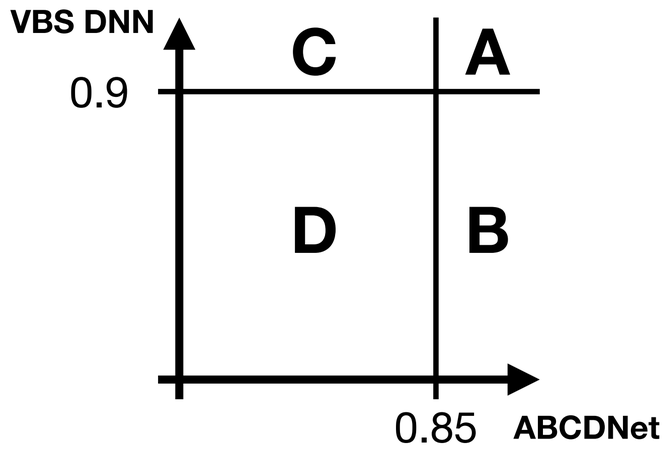
Xbb > 0.6 and XWqq > 0.75|0.7 (ld|tr)



Xbb > 0.8 and XWqq > 0.6|0.4 (ld|tr)



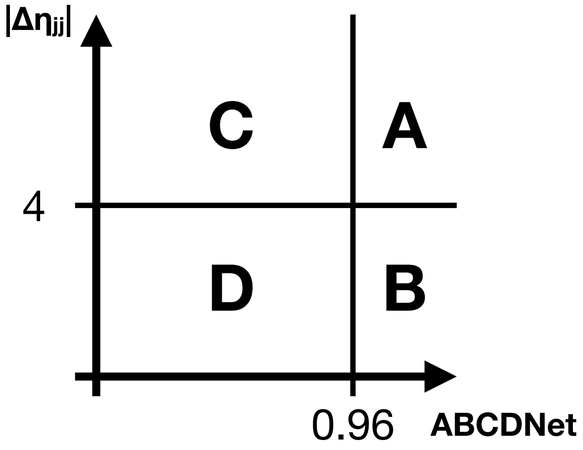
Xbb > 0.5 and XWqq > 0.5|0.4 (ld|tr)



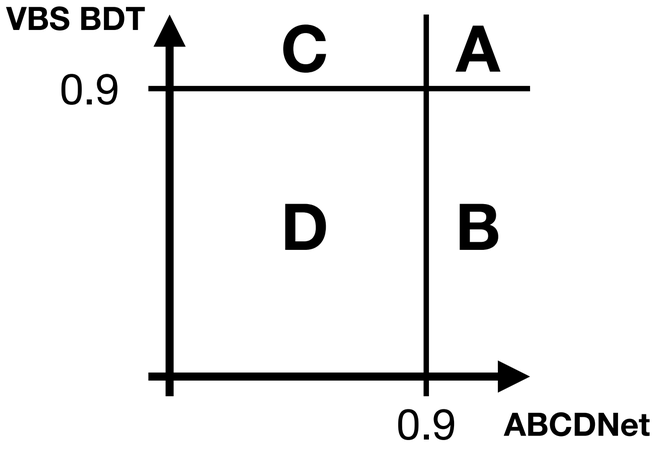
	ABCDNet vs. $\|\Delta\eta_{jj}\|$	ABCDNet vs. VBS BDT	ABCDNet vs. VBS DNN
Signal	3.73	3.80	6.88
Predicted Bkg.	0.93	1.09	0.78
ABCD Stat. Unc.	50.4%	18.9%	18.9%
ABCD Syst. Unc.	29.5%	62.9%	32.9%
Data/MC	OK	OK	Some issue from D→C
Closure in data	Good	Good, but limited by signal contamination	Good

Summary

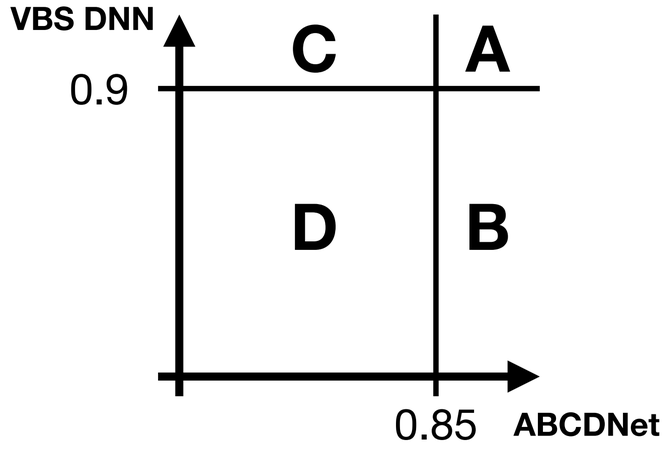
Xbb > 0.6 and XWqq > 0.75|0.7 (ld|tr)



Xbb > 0.8 and XWqq > 0.6|0.4 (ld|tr)



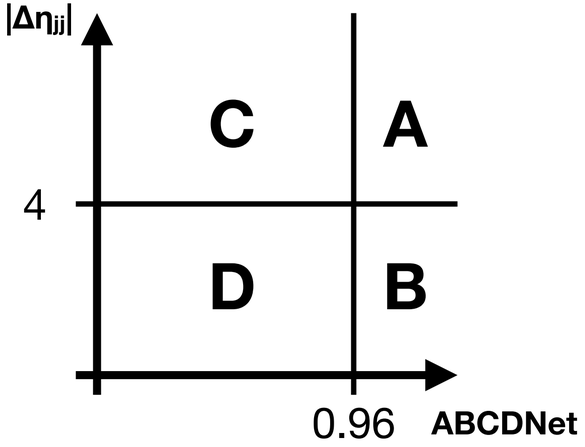
Xbb > 0.5 and XWqq > 0.5|0.4 (ld|tr)



- **No one “best” method**
- Systematic uncertainty may need rethinking (seems unfair?)
- VBS BDT method is nice, but sig. contamination limits data closure tests
- VBS DNN method is tempting… but do we believe it?

ABCD: ABCDNet vs. $|\Delta\eta_{jj}|$ (orig.)

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

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

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

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

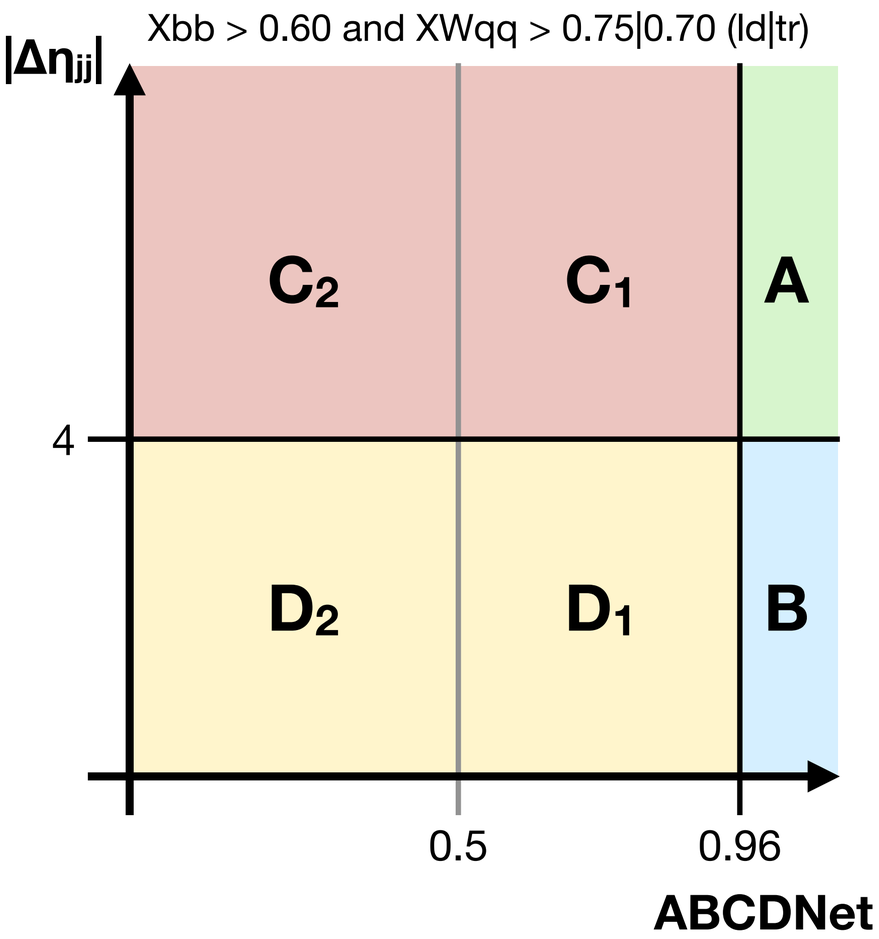
Predicted bkg. $0.93 \pm 0.47 \pm 0.28$

stat. *syst.*

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

UC San Diego

Closure in Sidebands



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

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

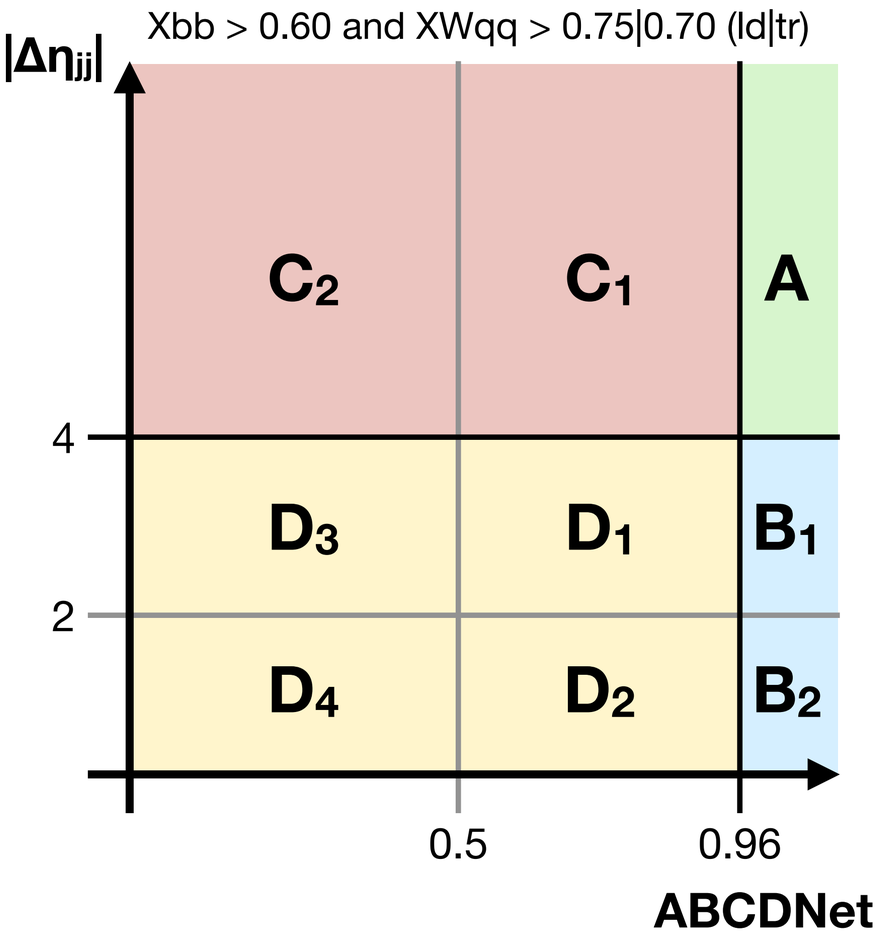
✓

$$A^{pred} = B \times \frac{C_1}{D_1} = \begin{bmatrix} \mathbf{1.08 \pm 0.30} & \text{(MC)} \\ ? \pm ? & \text{(Data)} \end{bmatrix}$$

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

UC San Diego

Closure in Sidebands



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

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

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

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

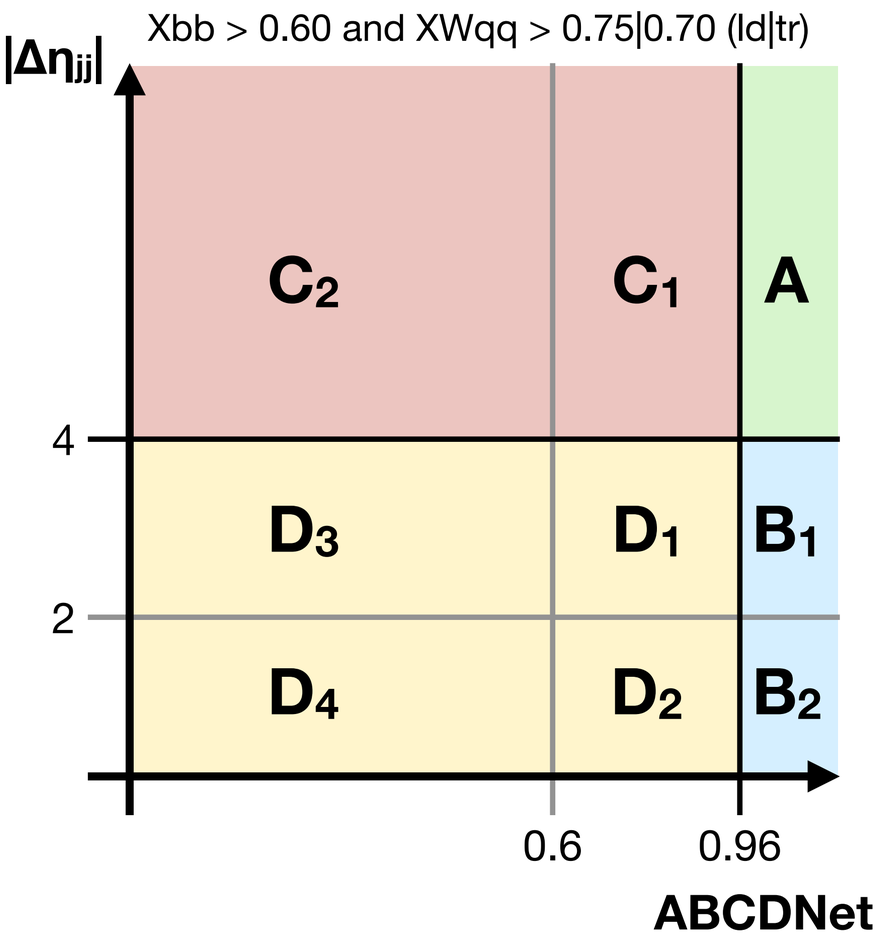
Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	17.17	2.46	2.29	0.04	–	–
D_1	33.28	3.80	0.20	0.01	37	6.08
C_2	275.71	25.78	0.51	0.02	264	16.25
D_3	441.29	28.47	0.07	0.01	540	23.24

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

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

UC San Diego

Closure in Sidebands



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

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

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

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

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

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

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

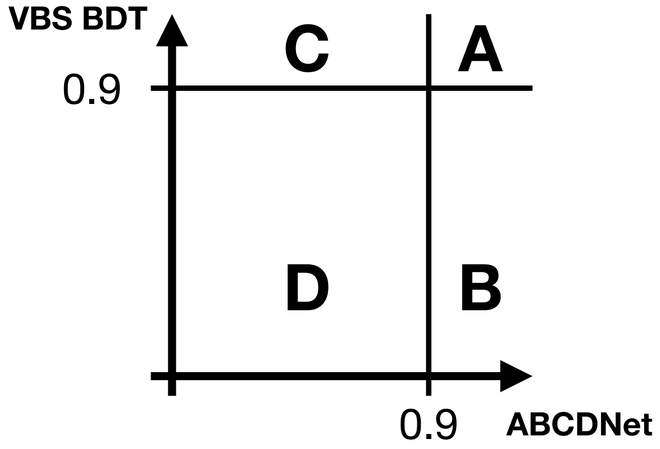
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{C1 \text{ is signal polluted} \Rightarrow \text{can't unblind}}$$

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

UC San Diego

ABCD: ABCDNet vs. VBS BDT

ABCD Configuration & MC Closure



Xbb > 0.8 and XWqq > 0.6|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	**0.72**	0.28	3.80	0.05	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	40.56	3.52	2.71	0.05	44	6.63
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	48.72	7.89	1.38	0.03	80	8.94
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	3025.85	79.24	1.30	0.03	3243	56.95

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{0.65 \pm 0.12}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{1.09 \pm 0.20}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{0.69}\sqrt{\left(\frac{0.28}{0.72}\right)^2 + \left(\frac{0.12}{0.65}\right)^2} = 62.9\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 18.9\%$$

Final Result

Expected sig. 3.80 ± 0.05

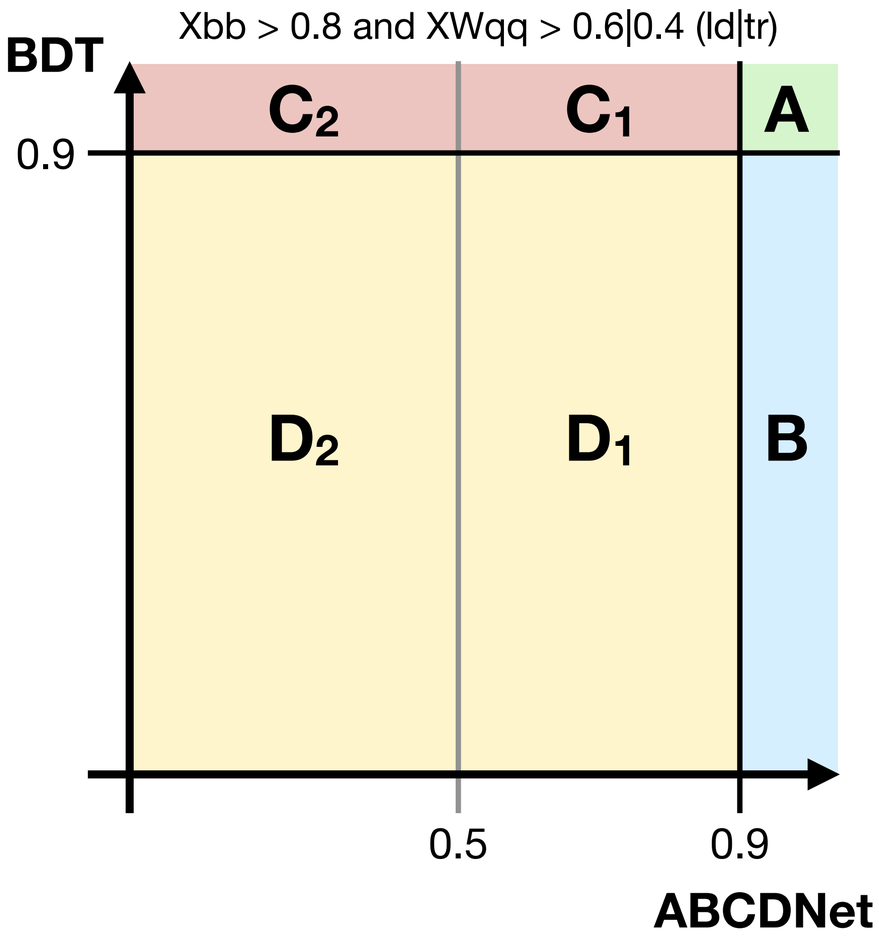
Predicted bkg. $1.09 \pm 0.20 \pm 0.68$

stat. *syst.*

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

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.72**	0.28	**3.80**	0.05	–	–
B	40.56	3.52	2.71	0.05	44	6.63
C	48.72	7.89	1.38	0.03	80	8.94
D	3025.8	79.24	1.30	0.03	3243	56.95

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} \mathbf{0.65 \pm 0.12} & \text{(MC)} \\ 1.09 \pm 0.20 & \text{(Data)} \end{bmatrix}$$

✓

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.72**	0.28	**3.80**	0.05	–	–
B	40.56	3.52	2.71	0.05	44	6.63
C_1	2.63	0.95	1.01	0.03	–	–
D_1	199.4	13.27	0.84	0.03	183	13.53

$$A^{pred} = B \times \frac{C_1}{D_1} = \begin{bmatrix} \mathbf{0.54 \pm 0.20} & \text{(MC)} \\ ? \pm ? & \text{(Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	2.63	0.95	1.01	0.03	–	–
D_1	199.43	13.27	0.84	0.03	183	13.53
C_2	46.09	7.83	0.37	0.02	76	8.72
D_2	2826.42	78.12	0.46	0.02	3060	55.32

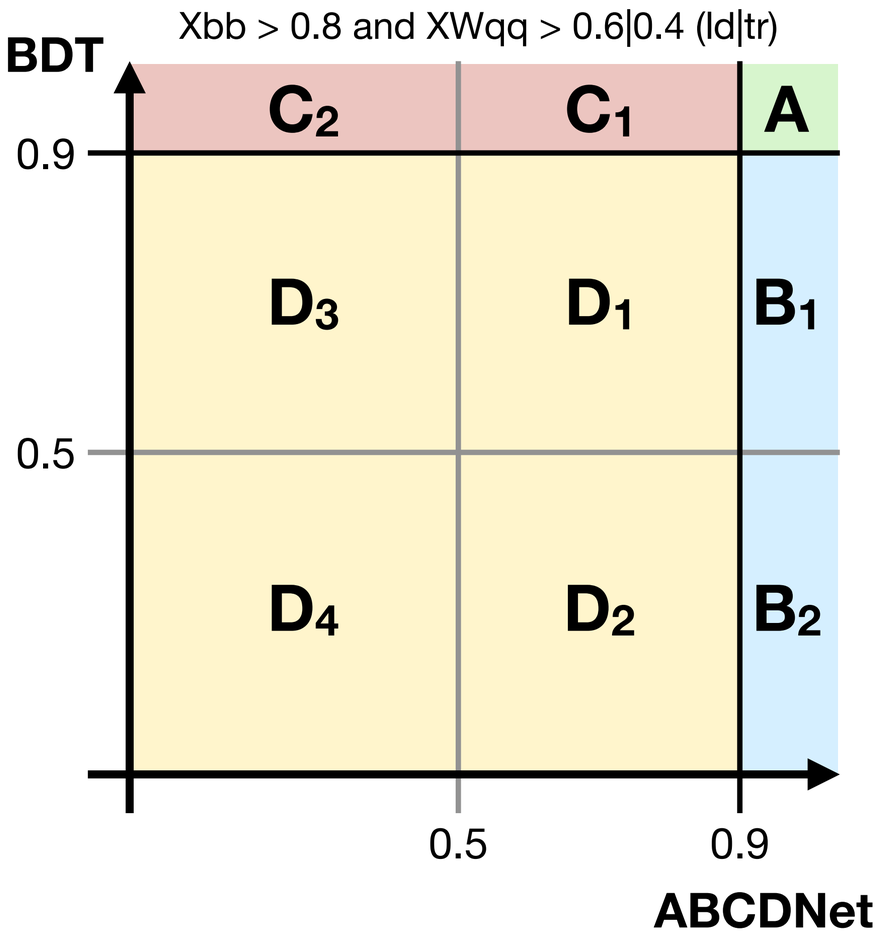
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{4.55 \pm 0.63}\ \text{(Data)}\ \mathbf{?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.72**	0.28	**3.80**	0.05	–	–
B	40.56	3.52	2.71	0.05	44	6.63
C	48.72	7.89	1.38	0.03	80	8.94
D	3025.85	79.24	1.30	0.03	3243	56.95

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} 0.65 \pm 0.12 \text{ (MC)} \\ \mathbf{1.09 \pm 0.20} \text{ (Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	4.94	1.06	1.92	0.04	–	–
B_2	35.62	3.36	0.79	0.02	39	6.24
D_1	18.94	2.93	0.56	0.02	21	4.58
D_2	180.48	12.95	0.28	0.01	162	12.73

$$B_1^{pred} = B_2 \times \frac{D_1}{D_2} = \mathbf{5.06 \pm 1.42} \text{ (Data) ?}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	18.94	2.93	0.56	0.02	**21**	4.58
D_2	180.48	12.95	0.28	0.01	162	12.73
D_3	279.77	20.77	0.24	0.01	320	17.89
D_4	2546.65	75.31	0.21	0.01	2740	52.35

$$D_1^{pred} = D_2 \times \frac{D_3}{D_4} = \mathbf{18.92 \pm 1.86} \text{ (Data) } \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	2.63	0.95	1.01	0.03	–	–
D_1	18.94	2.93	0.56	0.02	21	4.58
C_2	46.09	7.83	0.37	0.02	76	8.72
D_3	279.77	20.77	0.24	0.01	320	17.89

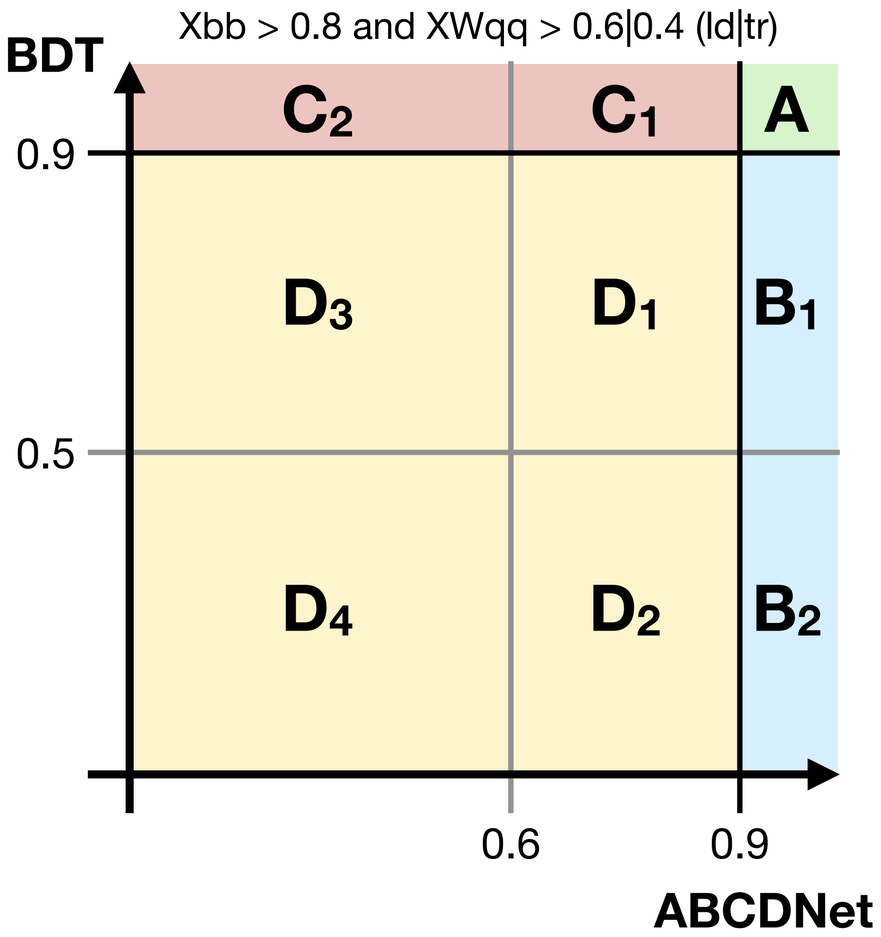
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{4.99 \pm 1.26} \text{ (Data) ?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.72**	0.28	**3.80**	0.05	–	–
B	40.56	3.52	2.71	0.05	44	6.63
C	48.72	7.89	1.38	0.03	80	8.94
D	3025.85	79.24	1.30	0.03	3243	56.95

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} 0.65 \pm 0.12 \text{ (MC)} \\ \mathbf{1.09 \pm 0.20} \text{ (Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	4.94	1.06	1.92	0.04	–	–
B_2	35.62	3.36	0.79	0.02	39	6.24
D_1	12.46	2.24	0.50	0.02	15	3.87
D_2	130.54	12.00	0.24	0.01	120	10.95

$$B_1^{pred} = B_2 \times \frac{D_1}{D_2} = \mathbf{4.88 \pm 1.55} \text{ (Data) ?}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	12.46	2.24	0.50	0.02	**15**	3.87
D_2	130.54	12.00	0.24	0.01	120	10.95
D_3	286.26	20.86	0.30	0.02	326	18.06
D_4	2596.59	75.47	0.25	0.01	2782	52.74

$$D_1^{pred} = D_2 \times \frac{D_3}{D_4} = \mathbf{14.06 \pm 1.52} \text{ (Data) } \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	2.09	0.88	0.91	0.03	–	–
D_1	12.46	2.24	0.50	0.02	15	3.87
C_2	46.63	7.84	0.47	0.02	77	8.77
D_3	286.26	20.86	0.30	0.02	326	18.06

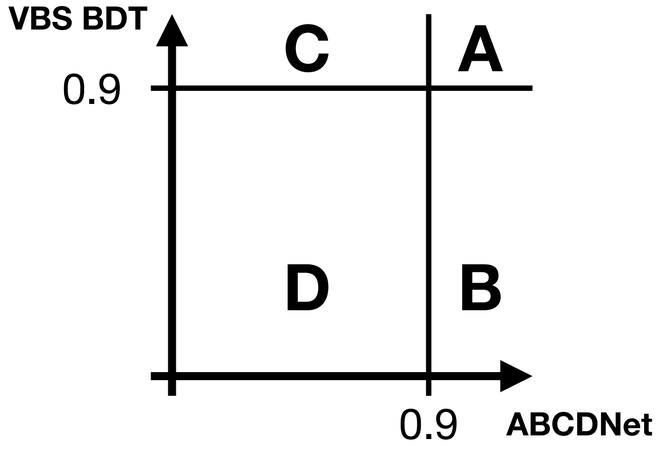
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{3.54 \pm 1.02} \text{ (Data) ?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

Closure in Xbb Sideband



Xbb ≤ 0.80 and XWqq > 0.60|0.40 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	**1.18**	0.64	0.29	0.01	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	39.35	3.99	0.25	0.01	49	7.0
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	49.98	10.74	0.27	0.01	58	7.6
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	2552.26	73.17	0.24	0.01	2746	52.4

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{0.77 \pm 0.18}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{1.03 \pm 0.20}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{0.97}\sqrt{\left(\frac{0.64}{1.18}\right)^2 + \left(\frac{0.18}{0.77}\right)^2} = 61.1\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 19.5\%$$

Final Result

Expected sig. 0.29 ± 0.01

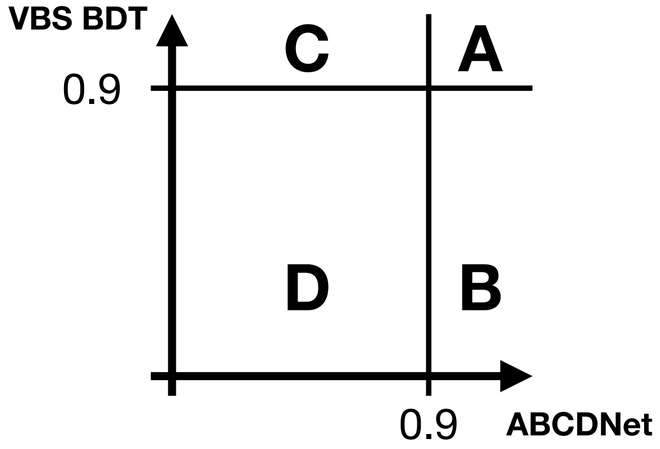
Predicted bkg. 1.03 ± 0.20 ± 0.63

stat. *syst.*

Method closes for data in ParticleNet sideband

UC San Diego

Closure in ParticleNet Sideband



Xbb ≤ 0.80 or XWqq ≤ 0.60|0.40 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	**4.02**	1.19	0.75	0.02	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	155.80	7.75	0.59	0.02	155	12.4
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	237.20	24.86	0.55	0.02	211	14.5
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	10832.14	162.63	0.64	0.02	10609	103.0

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{3.41 \pm 0.40}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{3.08 \pm 0.33}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{3.71}\sqrt{\left(\frac{1.19}{4.02}\right)^2 + \left(\frac{0.40}{3.41}\right)^2} = 8.6\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 10.6\%$$

Final Result

Expected sig. 0.75 ± 0.02

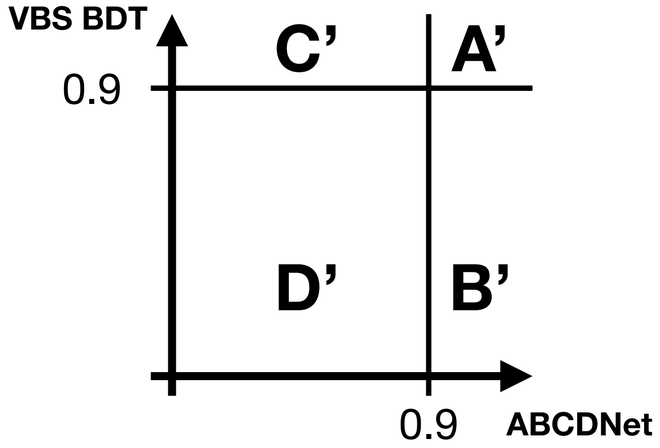
Predicted bkg. 3.08 ± 0.33 ± 0.26

stat. *syst.*

Method closes for data in ParticleNet sideband

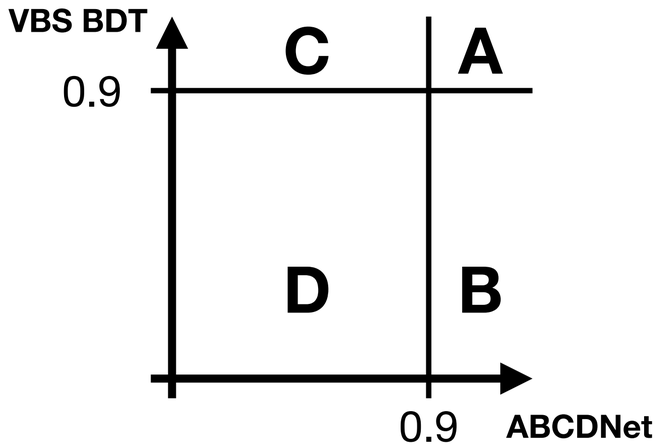
UC San Diego

Closure in ParticleNet Sideband



Xbb ≤ 0.80 or XWqq ≤ 0.60|0.40 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A’	4.02	1.19	0.75	0.02	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B’	155.80	7.75	0.59	0.02	155	12.4
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C’	237.20	24.86	0.55	0.02	211	14.5
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D’	10832.14	162.63	0.64	0.02	10609	103.0



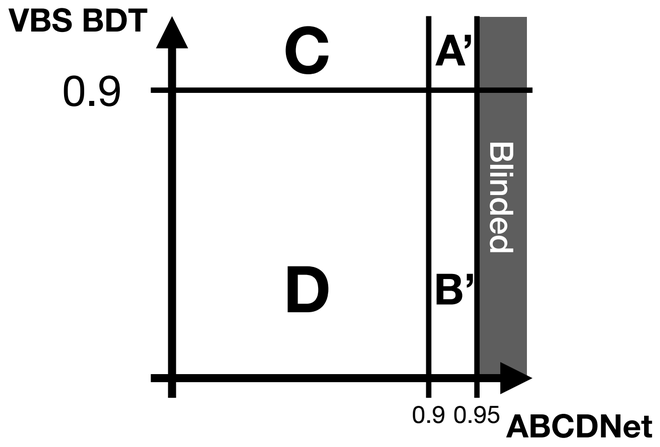
Xbb > 0.8 and XWqq > 0.6|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	0.72	0.28	3.80	0.05	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	40.56	3.52	2.71	0.05	44	6.63
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	48.72	7.89	1.38	0.03	80	8.94
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	3025.85	79.24	1.30	0.03	3243	56.95

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A’/A	5.54	2.72	5.08	0.18	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B’/B	3.84	0.38	4.58	0.18	3.5	0.60
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C’/C	4.87	0.94	2.51	0.11	2.6	0.35
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D’/D	3.58	0.11	2.03	0.09	3.3	0.07

UC San Diego

Closure in PNet + ABCDNet Sideband



Xbb ≤ 0.80 or XWqq ≤ 0.60|0.40 (Id|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A'	**2.05**	0.92	0.16	0.01	1	1.00
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B'	83.00	5.84	0.16	0.01	79	8.9
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	237.20	24.86	0.55	0.02	211	14.5
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	10832.14	162.63	0.64	0.02	10609	103.0

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{1.82 \pm 0.23}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{1.57 \pm 0.21}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{1.93}\sqrt{\left(\frac{0.92}{2.05}\right)^2 + \left(\frac{0.23}{1.82}\right)^2} = 24.3\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 13.2\%$$

Final Result

Expected sig. 0.16 ± 0.01

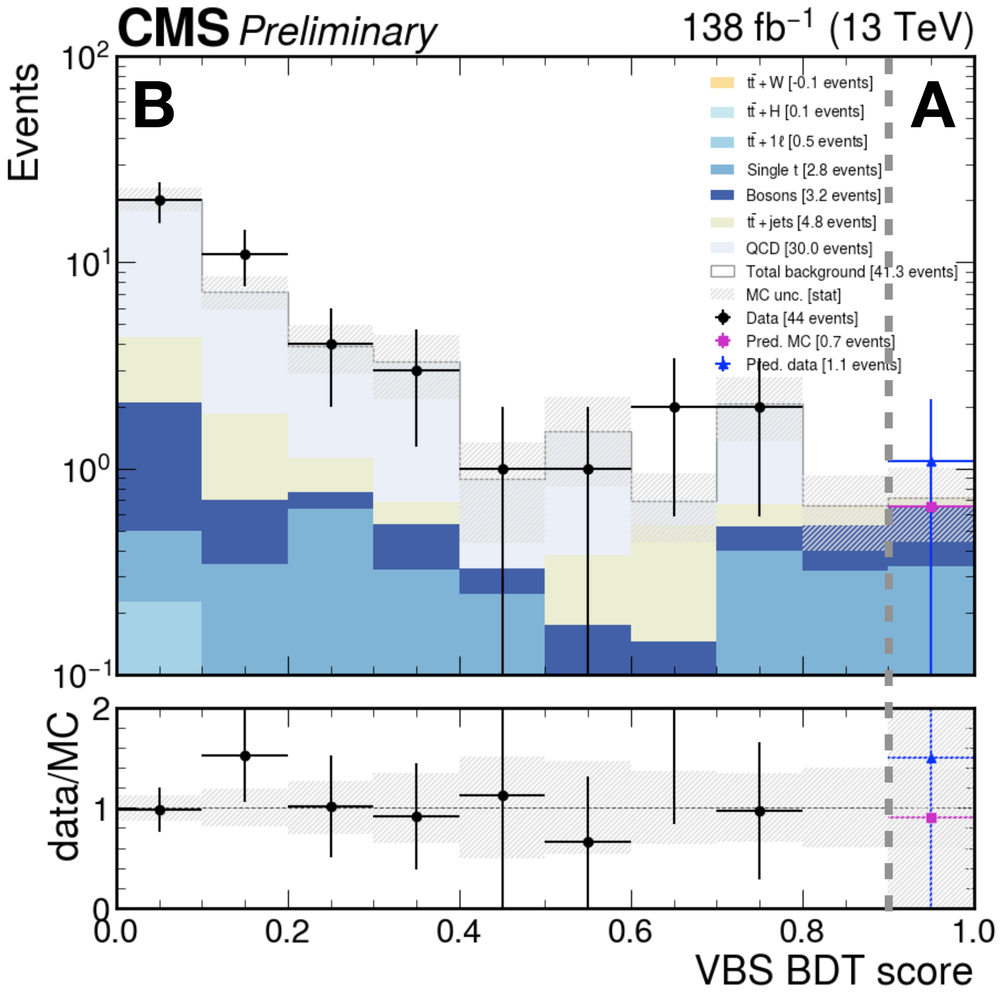
Predicted bkg. 1.57 ± 0.21 ± 0.38

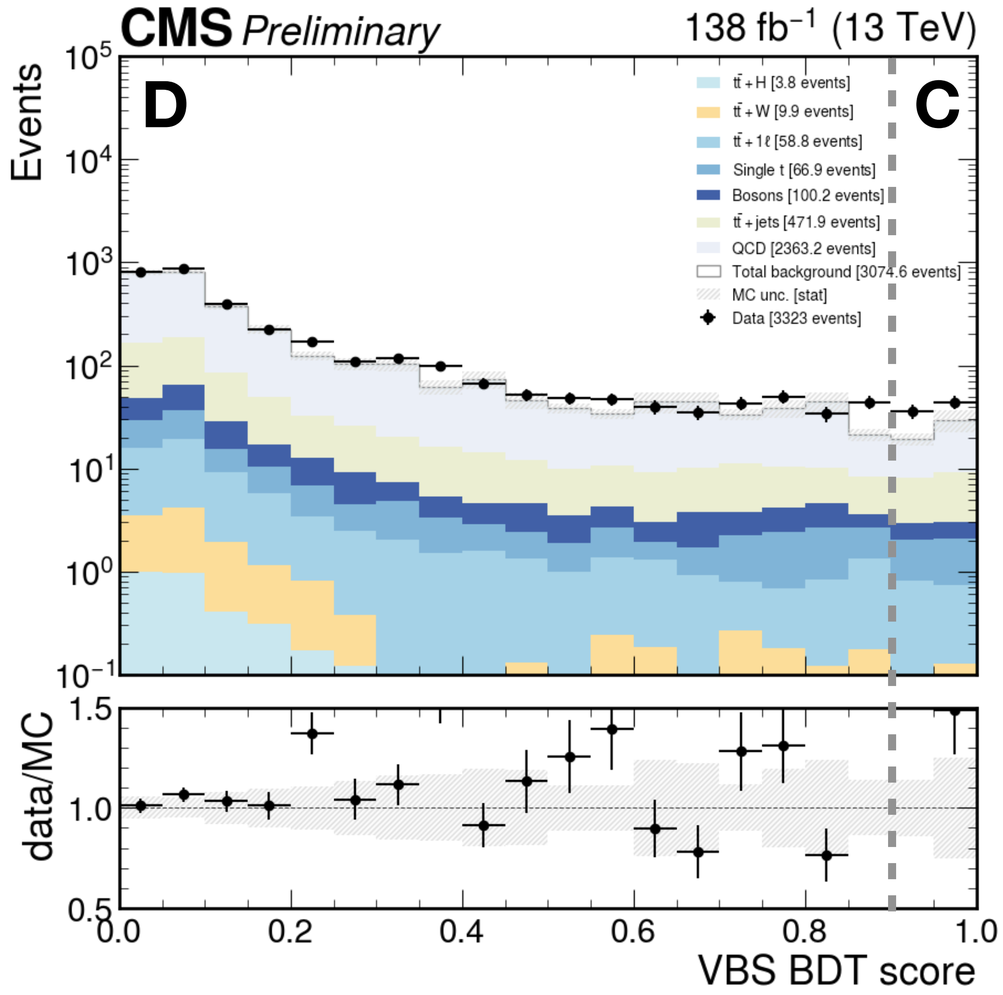
stat. *syst.*

Method closes for data in ParticleNet sideband

UC San Diego

Reweighted vs. Single-point



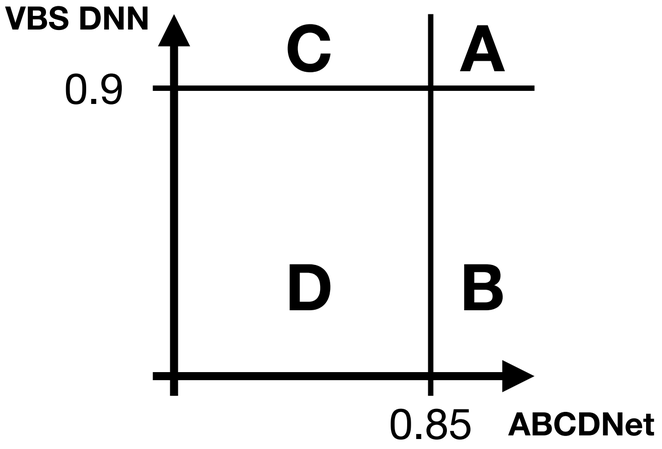


No obvious systematic issues in data/MC

UC San Diego

ABCD: ABCDNet vs. VBS DNN

ABCD Configuration & MC Closure



Xbb > 0.5 and XWqq > 0.5|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS DNN > 0.9 and ABCDNet > 0.9	A	**0.95**	0.37	6.88	0.07	—	—
VBS DNN ≤ 0.9 and ABCDNet > 0.9	B	189.89	8.37	0.92	0.03	185	13.60
VBS DNN > 0.9 and ABCDNet ≤ 0.9	C	56.11	2.29	2.35	0.04	33	5.74
VBS DNN ≤ 0.9 and ABCDNet ≤ 0.9	D	7271.69	125.39	0.54	0.02	7817	88.41

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{1.47 \pm 0.09}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{0.78 \pm 0.15}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{1.21}\sqrt{\left(\frac{0.37}{0.95}\right)^2 + \left(\frac{0.09}{1.47}\right)^2} = 32.9\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 18.9\%$$

Final Result

Expected sig. 6.88 ± 0.07

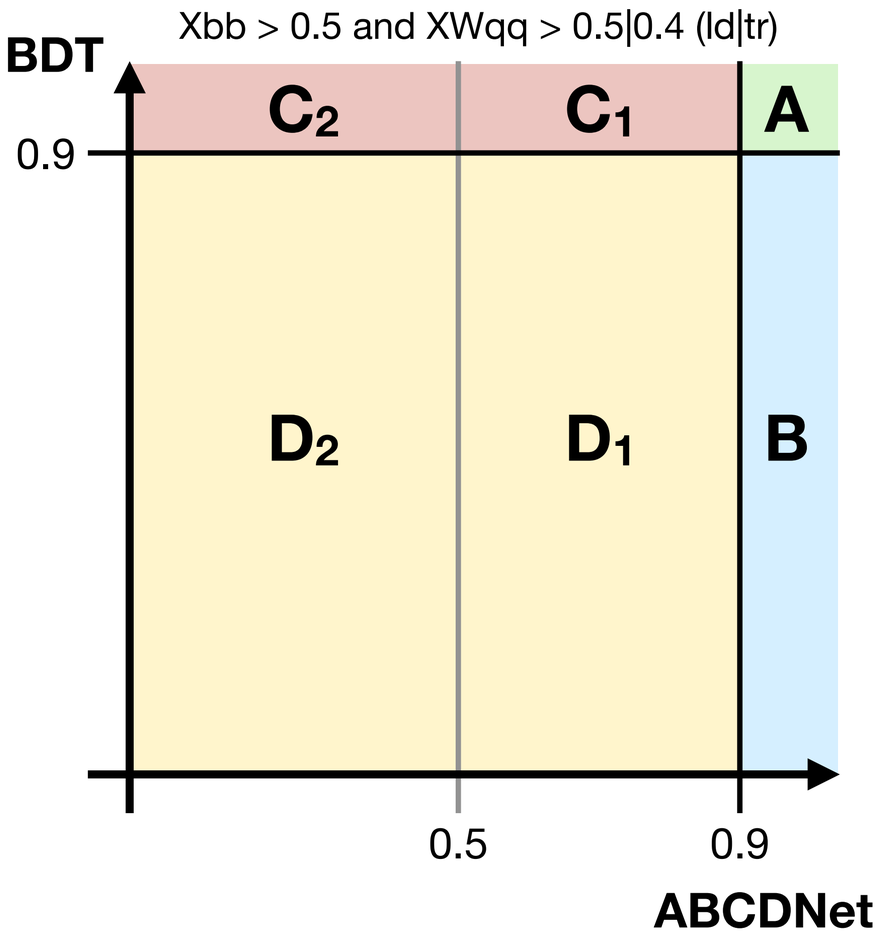
Predicted bkg. $0.78 \pm 0.15 \pm 0.26$

stat. *syst.*

Predicted significance (S/√B) is very good and the method roughly closes in MC

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.95**	0.37	**6.88**	0.07	–	–
B	189.89	8.37	0.92	0.03	185	13.60
C	56.11	2.29	2.35	0.04	33	5.74
D	7271.7	125.39	0.54	0.02	7817	88.41

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} \mathbf{1.47 \pm 0.09} & \text{(MC)} \\ 0.78 \pm 0.15 & \text{(Data)} \end{bmatrix}$$

~

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.95**	0.37	**6.88**	0.07	–	–
B	189.89	8.37	0.92	0.03	185	13.60
C_1	3.18	0.48	1.52	0.03	–	–
D_1	570.3	24.36	0.27	0.01	552	23.49

$$A^{pred} = B \times \frac{C_1}{D_1} = \begin{bmatrix} \mathbf{1.06 \pm 0.17} & \text{(MC)} \\ ? \pm ? & \text{(Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	3.18	0.48	1.52	0.03	–	–
D_1	570.27	24.36	0.27	0.01	552	23.49
C_2	52.94	2.23	0.83	0.02	31	5.57
D_2	6701.42	123.00	0.27	0.01	7265	85.23

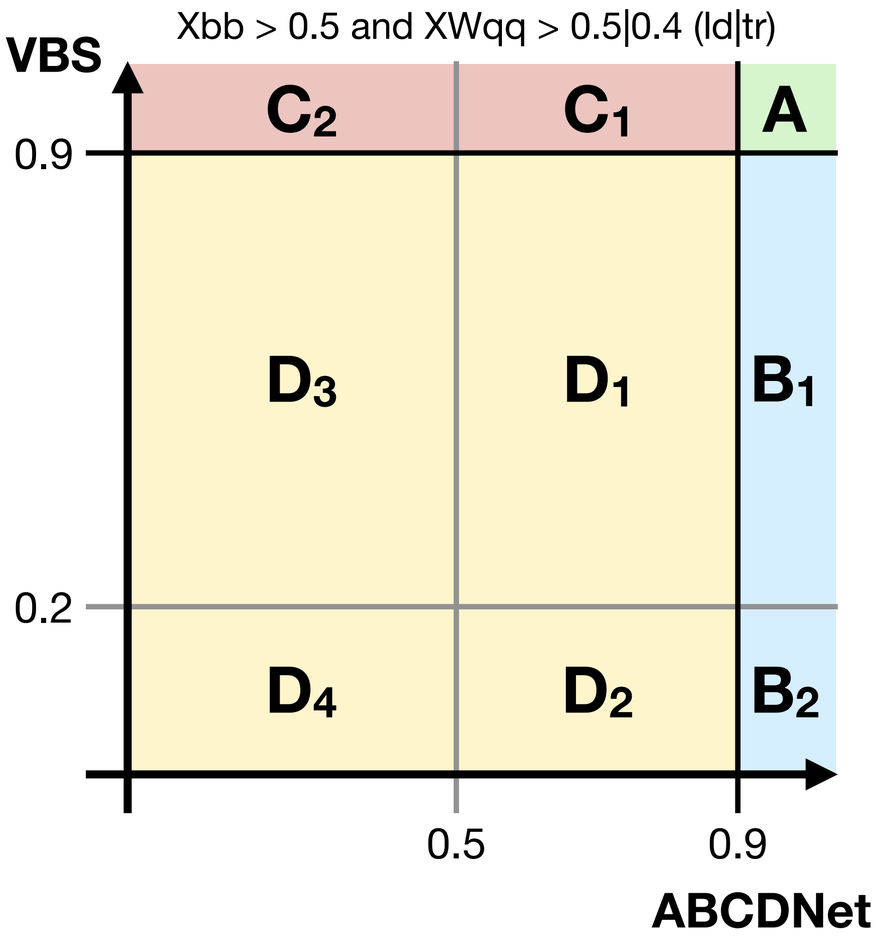
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{2.36 \pm 0.44} \text{ (Data) } \mathbf{?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.95**	0.37	**6.88**	0.07	–	–
B	189.89	8.37	0.92	0.03	185	13.60
C	56.11	2.29	2.35	0.04	33	5.74
D	7271.69	125.39	0.54	0.02	7817	88.41

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} 1.47 \pm 0.09 \text{ (MC)} \\ \mathbf{0.78 \pm 0.15} \text{ (Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	11.22	1.33	0.87	0.03	7	2.65
B_2	178.66	8.26	0.05	0.01	178	13.34
D_1	42.61	3.12	0.26	0.01	29	5.39
D_2	527.67	24.15	0.01	0.00	523	22.87

$$B_1^{pred} = B_2 \times \frac{D_1}{D_2} = \mathbf{9.87 \pm 2.02} \text{ (Data)} \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	42.61	3.12	0.26	0.01	29	5.39
D_2	527.67	24.15	0.01	0.00	523	22.87
D_3	514.84	14.30	0.25	0.01	380	19.49
D_4	6186.57	122.16	0.02	0.00	6885	82.98

$$D_1^{pred} = D_2 \times \frac{D_3}{D_4} = \mathbf{28.87 \pm 1.98} \text{ (Data)} \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	3.18	0.48	1.52	0.03	–	–
D_1	42.61	3.12	0.26	0.01	29	5.39
C_2	52.94	2.23	0.83	0.02	31	5.57
D_3	514.84	14.30	0.25	0.01	380	19.49

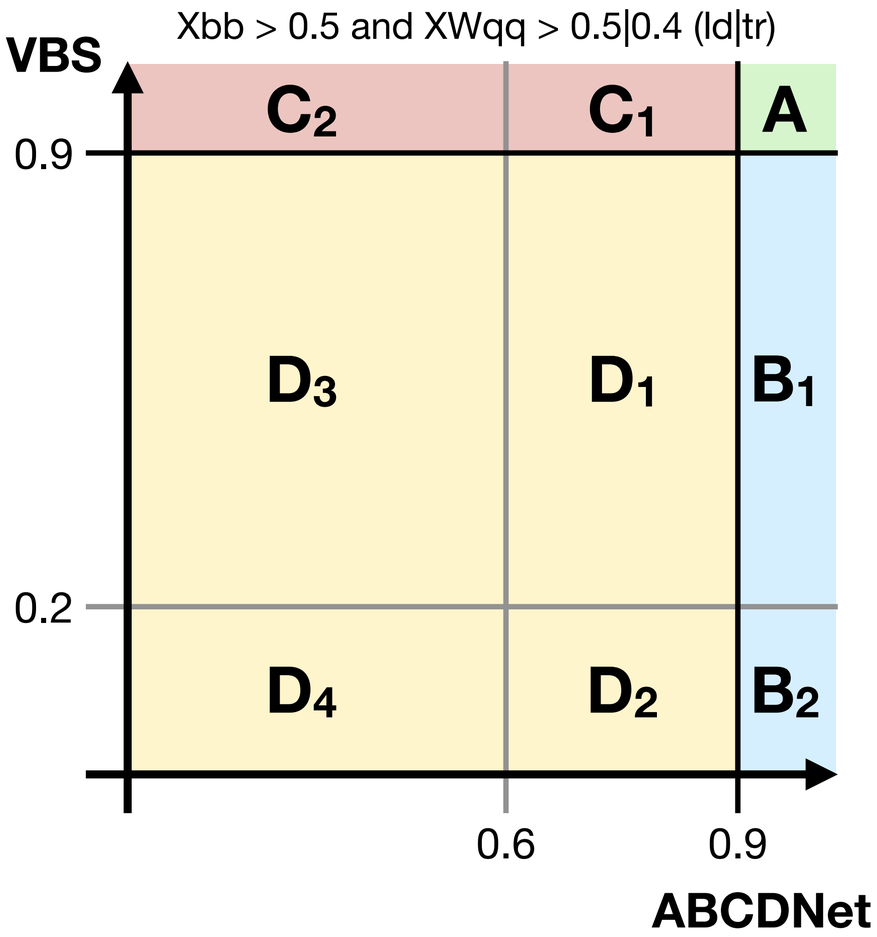
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{2.37 \pm 0.62} \text{ (Data)} \textbf{ ?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

Closure in Sidebands



Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
A	**0.95**	0.37	**6.88**	0.07	–	–
B	189.89	8.37	0.92	0.03	185	13.60
C	56.11	2.29	2.35	0.04	33	5.74
D	7271.69	125.39	0.54	0.02	7817	88.41

$$A^{pred} = B \times \frac{C}{D} = \begin{bmatrix} 1.47 \pm 0.09 \text{ (MC)} \\ \mathbf{0.78 \pm 0.15} \text{ (Data)} \end{bmatrix}$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
B_1	11.22	1.33	0.87	0.03	7	2.65
B_2	178.66	8.26	0.05	0.01	178	13.34
D_1	28.31	2.45	0.21	0.01	19	4.36
D_2	333.99	14.65	0.01	0.00	355	18.84

$$B_1^{pred} = B_2 \times \frac{D_1}{D_2} = \mathbf{9.53 \pm 2.35} \text{ (Data)} \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
D_1	28.31	2.45	0.21	0.01	19	4.36
D_2	333.99	14.65	0.01	0.00	355	18.84
D_3	529.14	14.43	0.30	0.02	390	19.75
D_4	6380.25	123.66	0.02	0.00	7053	83.98

$$D_1^{pred} = D_2 \times \frac{D_3}{D_4} = \mathbf{19.63 \pm 1.46} \text{ (Data)} \checkmark$$

Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
C_1	2.42	0.43	1.29	0.03	–	–
D_1	28.31	2.45	0.21	0.01	19	4.36
C_2	53.69	2.24	1.05	0.03	31	5.57
D_3	529.14	14.43	0.30	0.02	390	19.75

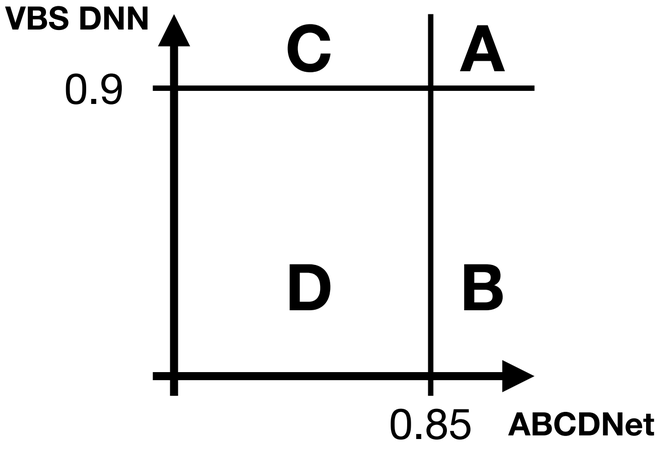
$$C_1^{pred} = D_1 \times \frac{C_2}{D_3} = \mathbf{1.51 \pm 0.45} \text{ (Data) ?}$$

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

Epoch = 600 | LR = 0.001 (constant) | λ = 30 | QCD norm | All features normalized

UC San Diego

ABCD Configuration & MC Closure



Xbb > 0.5 and XWqq ≤ 0.5|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	**0.71**	0.27	0.53	0.02	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	177.14	10.12	0.09	0.01	168	13.0
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	30.90	1.54	0.30	0.02	15	3.9
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	6617.60	132.59	0.10	0.01	6125	78.3

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{0.83 \pm 0.06}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{0.41 \pm 0.11}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{0.77}\sqrt{\left(\frac{0.27}{0.71}\right)^2 + \left(\frac{0.06}{0.83}\right)^2} = 49.9\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 27.0\%$$

Final Result

Expected sig. 0.53 ± 0.02

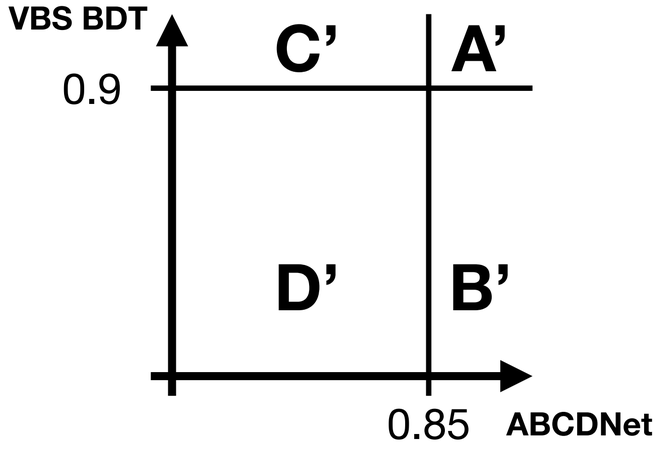
Predicted bkg. $0.41 \pm 0.11 \pm 0.21$

stat. *syst.*

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

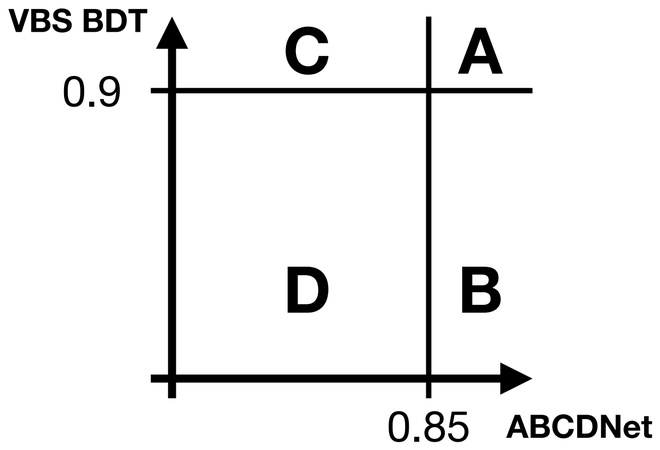
UC San Diego

Closure in ParticleNet Sideband



Xbb > 0.5 and XWqq ≤ 0.5|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A’	0.71	0.27	0.53	0.02	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B’	177.14	10.12	0.09	0.01	168	13.0
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C’	30.90	1.54	0.30	0.02	15	3.9
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D’	6617.60	132.59	0.10	0.01	6125	78.3



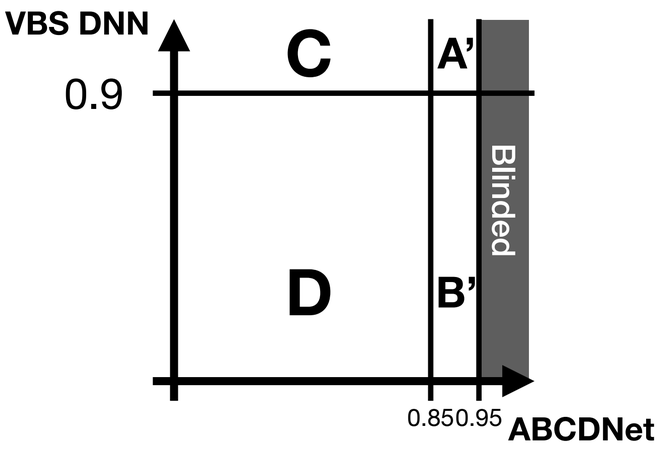
Xbb > 0.5 and XWqq > 0.5|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A	0.95	0.37	6.88	0.07	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B	189.89	8.37	0.92	0.03	185	13.6
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	56.11	2.29	2.35	0.04	33	5.7
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	7271.69	125.39	0.54	0.02	7817	88.4

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A’/A	0.74	0.40	12.97	0.51	—	—
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B’/B	0.93	0.07	10.01	0.96	0.9	0.10
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C’/C	0.55	0.04	7.71	0.41	0.5	0.14
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D’/D	0.91	0.02	5.59	0.55	0.8	0.01

UC San Diego

ABCD Configuration & MC Closure



Xbb > 0.5 and XWqq ≤ 0.5|0.4 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS BDT > 0.9 and ABCDNet > 0.9	A'	**0.43**	0.16	0.18	0.01	1	1.0
VBS BDT ≤ 0.9 and ABCDNet > 0.9	B'	125.47	9.36	0.04	0.01	133	11.53
VBS BDT > 0.9 and ABCDNet ≤ 0.9	C	30.90	1.54	0.30	0.02	15	3.87
VBS BDT ≤ 0.9 and ABCDNet ≤ 0.9	D	6617.60	132.59	0.10	0.01	6125	78.26

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{0.59 \pm 0.05}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{0.33 \pm 0.09}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{0.51}\sqrt{\left(\frac{0.16}{0.43}\right)^2 + \left(\frac{0.05}{0.59}\right)^2} = 73.8\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 27.3\%$$

Final Result

Expected sig. 0.18 ± 0.01

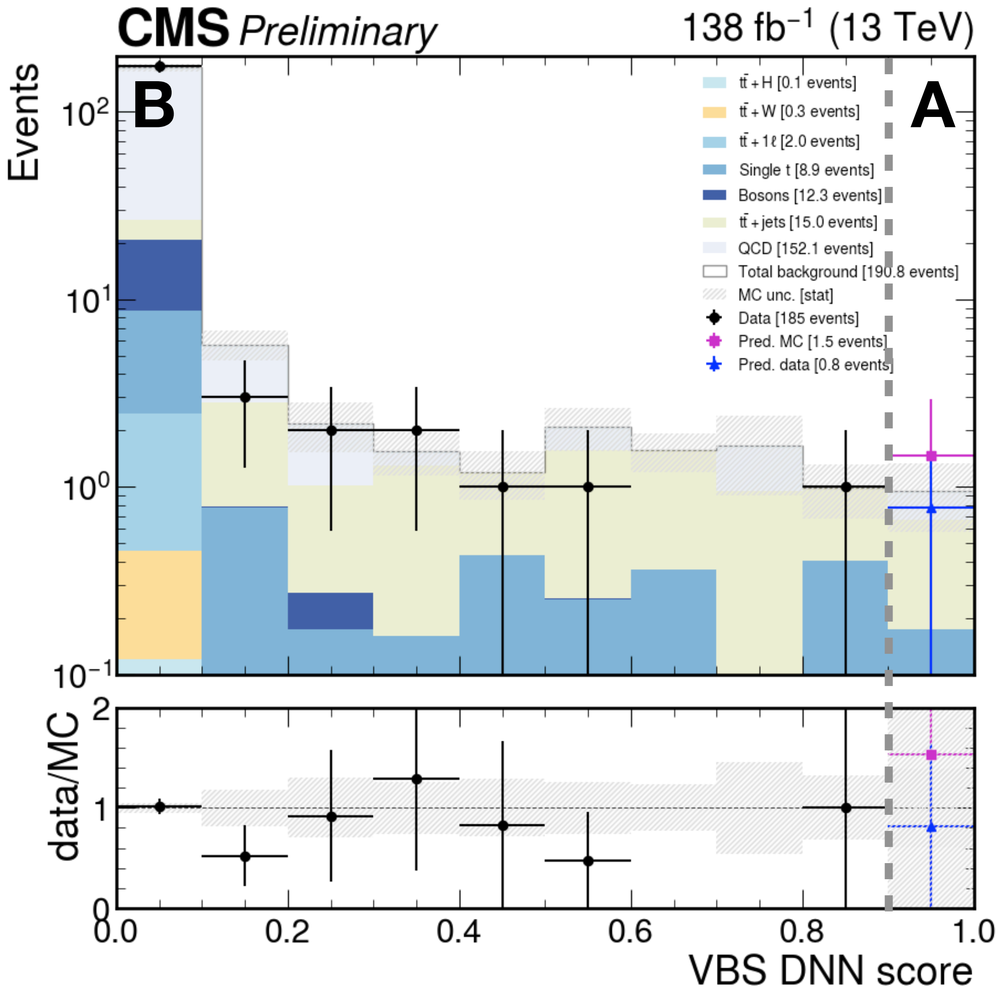
Predicted bkg. $0.33 \pm 0.09 \pm 0.24$

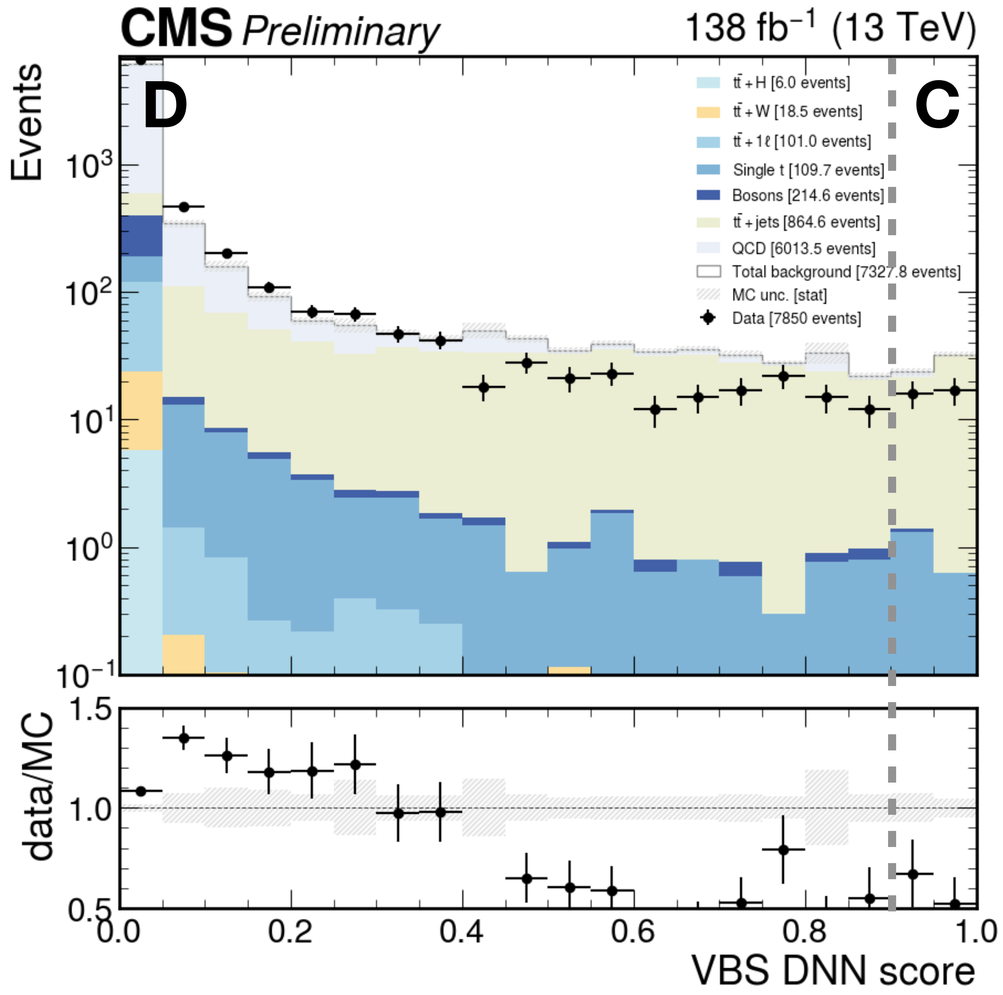
stat. *syst.*

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

UC San Diego

Reweighted vs. Single-point



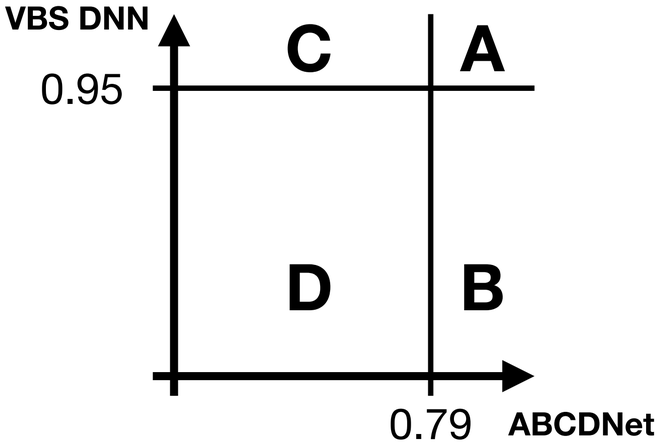


Seems to be a systematic issue in data/MC going from D to C

UC San Diego

Backup

ABCD Configuration & MC Closure



Xbb > 0.5 and XWqq > 0.3|0.3 (ld|tr)

Selection	Region	Bkg	Bkg err	Sig	Sig err	Data	Data err
VBS DNN > 0.9 and ABCDNet > 0.9	A	**1.00**	0.28	7.23	0.08	—	—
VBS DNN ≤ 0.9 and ABCDNet > 0.9	B	532.17	15.82	1.86	0.04	510	22.58
VBS DNN > 0.9 and ABCDNet ≤ 0.9	C	52.21	1.96	1.83	0.04	23	4.80
VBS DNN ≤ 0.9 and ABCDNet ≤ 0.9	D	13759.61	182.28	0.79	0.02	13811	117.52

$$A_{MC} = B_{MC} \times C_{MC} / D_{MC} = \mathbf{2.02 \pm 0.10}$$

$$A_{pred} = B_{data} \times C_{data} / D_{data} = \mathbf{0.85 \pm 0.18}$$

Error propagation:

$$\epsilon_{syst} = \frac{1}{1.51}\sqrt{\left(\frac{0.28}{1.00}\right)^2 + \left(\frac{0.10}{2.02}\right)^2} = 19.0\%$$

$$\epsilon_{stat} = \sqrt{\frac{1}{B_{data}} + \frac{1}{C_{data}} + \frac{1}{D_{data}}} = 21.3\%$$

Final Result

Expected sig. 7.23 ± 0.08

Predicted bkg. $0.85 \pm 0.18 \pm 0.16$

stat. *syst.*

Predicted significance (S/√B) is very good and the method roughly closes in MC

UC San Diego