VBS WH Analysis

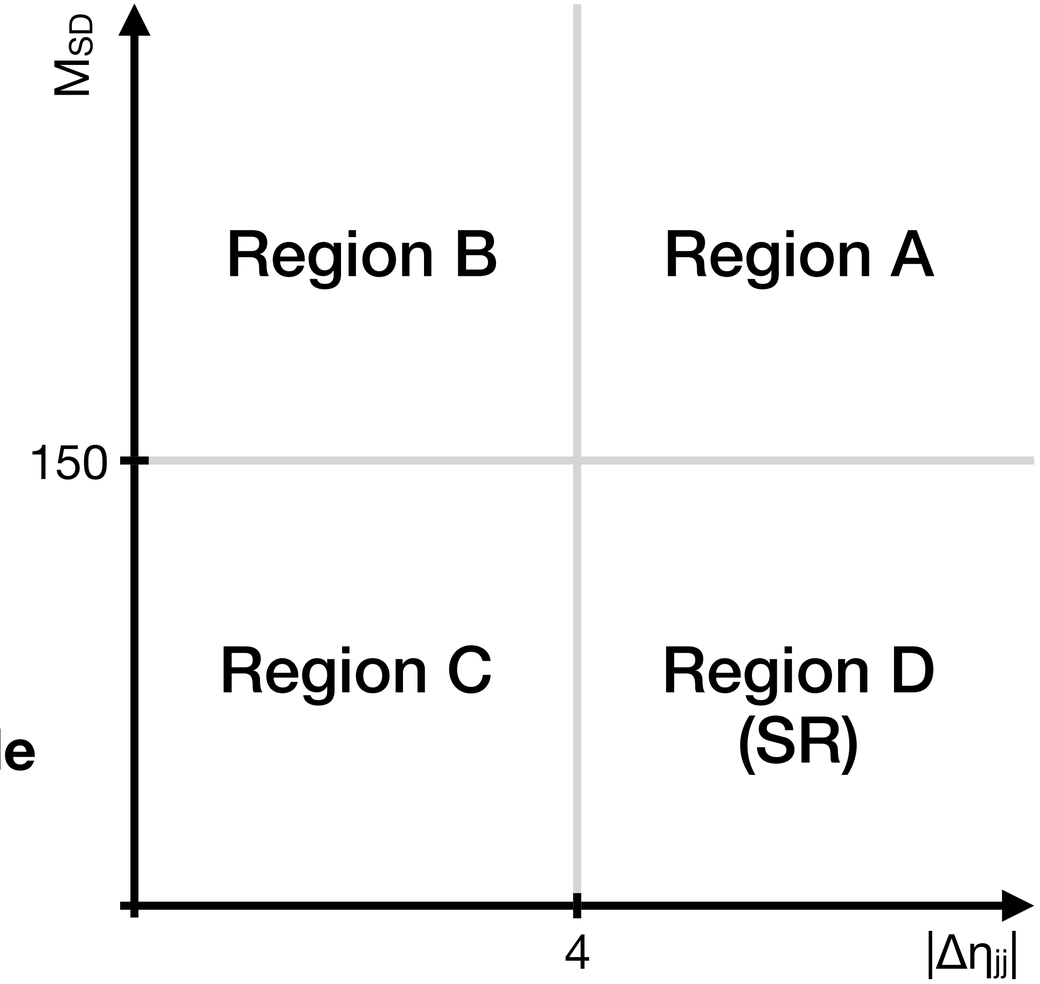
Proposed ABCD correction

June 16th, 2023

*P. Chang, L. Giannini, **J. Guiang**, F. Würthwein*

Overview

- We were asked to do two closure tests in data:
  - ABCD in a ParticleNet Xbb sideband*
  - ABCD in Regions A+B (i.e. $M_{SD} \geq 150$ GeV)
- We were also asked to make correlation plots
- The results are presented here
- **We find that the method is not perfect, but with some correction, it provides a reasonable background estimate that is precise enough for our analysis (S/B ~ 360/100)**



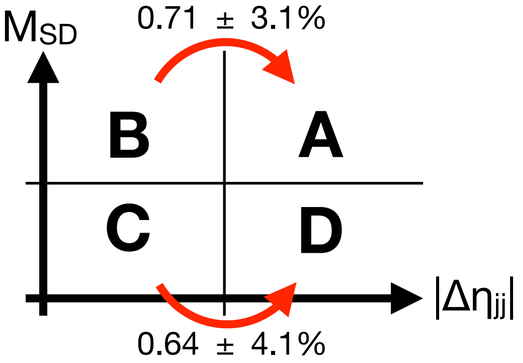
*Any sideband near enough to the SR is too signal-polluted

Recall: Original ABCD Method

Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Total Bkg.	Sig. (λ_{WZ} = -1)	Data
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} ≥ 150 GeV	A	173.0 ± 3.3	6.9 ± 3.3	142 ± 11.9
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	241.9 ± 5.8	0.3 ± 5.8	201 ± 14.2
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} < 150 GeV	C	181.1 ± 4.4	11.6 ± 4.4	170 ± 13.0
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} < 150 GeV (SR)	D	116.4 ± 3.8	366.3 ± 3.8	—

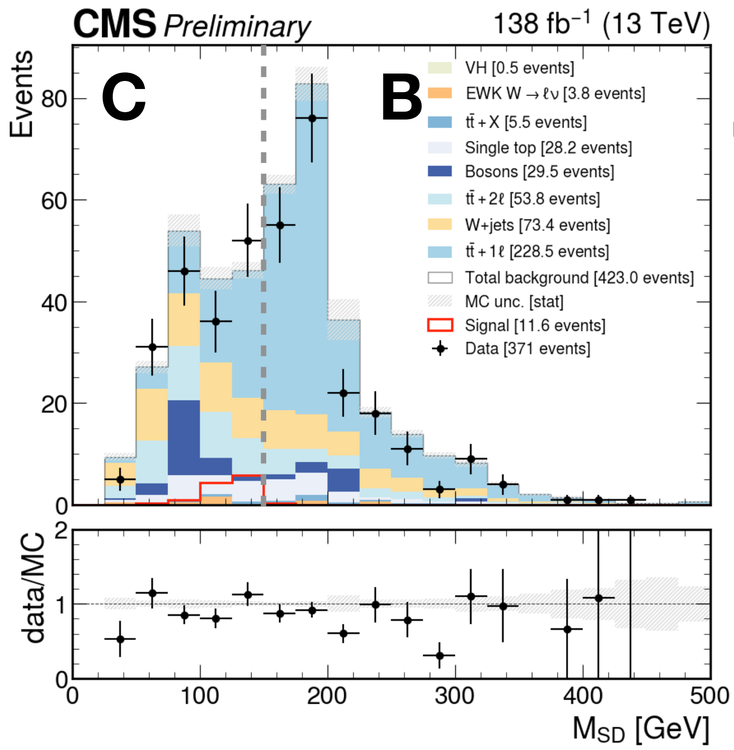
*



- Data an MC agree reasonably well in B, C, D
  - Only use yields in final limit
- Data and MC transfer factors agree

$$\frac{C_{MC}}{D_{MC}} = 0.71 \pm 3.1\%$$

$$\frac{C_{data}}{D_{data}} = 0.71 \pm 11.0\%$$



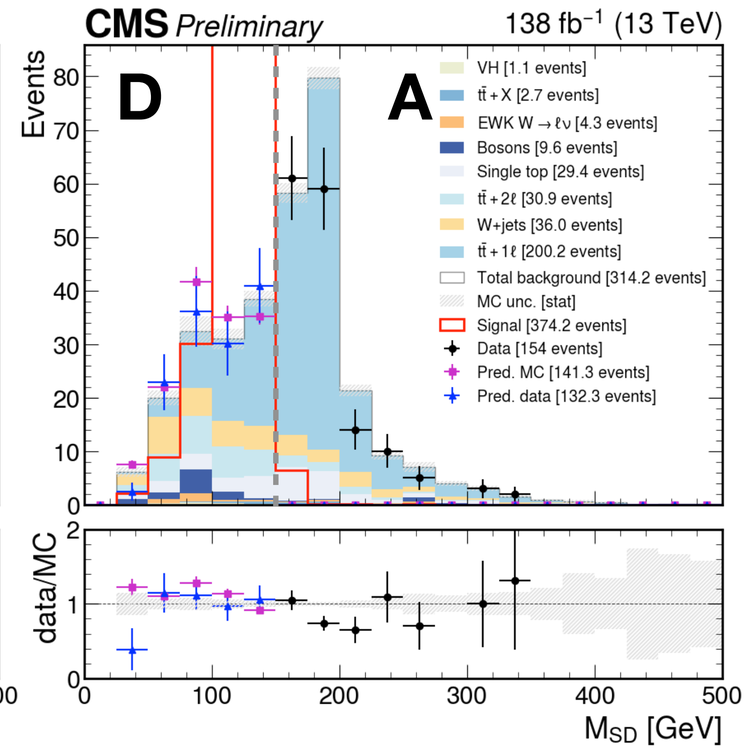


*Table errors = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

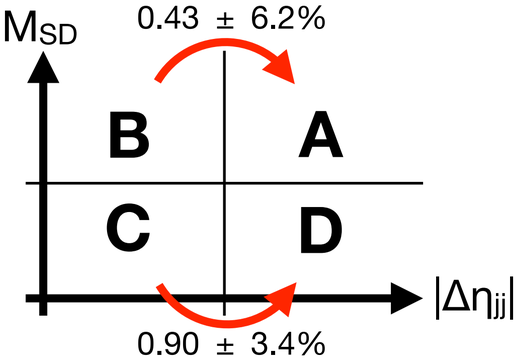
UC San Diego

ABCD in M_{SD} Sideband

Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 AND M_{SD} ≥ 150 GeV

Cut	Region	Total Bkg.	Sig. (λ_{WZ} = -1)	Data
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} ≥ 200 GeV	A	41.8 ± 1.7	0.3 ± 1.7	28 ± 5.3
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND $M_{SD}\in$ ≥ 200 GeV	B	96.1 ± 4.4	0.0 ± 4.4	70 ± 8.4
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} < 200 GeV	C	145.9 ± 3.8	0.2 ± 3.8	131 ± 11.5
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} < 200 GeV ("SR")	D	131.2 ± 2.8	6.6 ± 2.8	**114 ± 10.7**

*

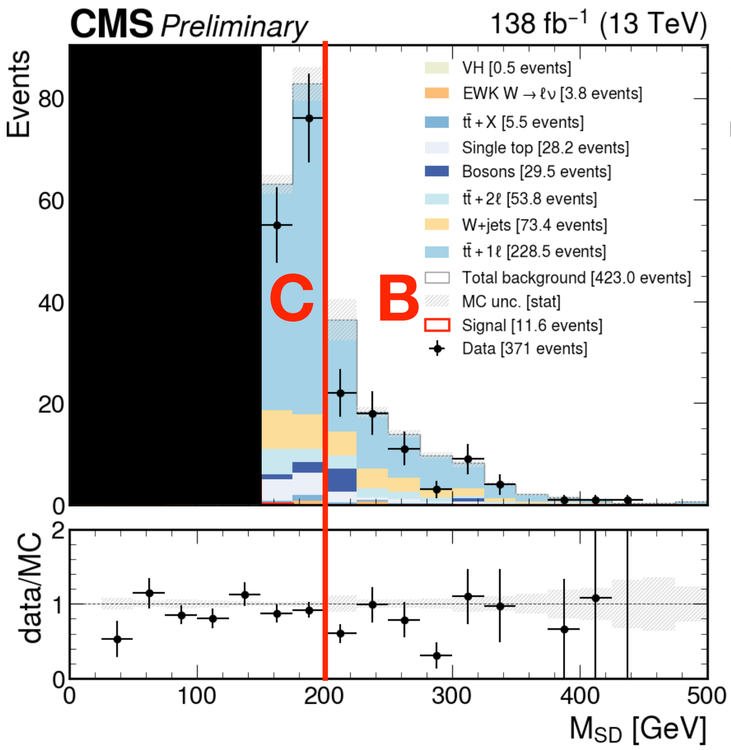


- Closure in data is **very bad**:

$$D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}} = \boxed{52.4 \pm 12.6}$$

- If M_{SD} and $\Delta\eta_{jj}$ are decorrelated, ABCD should work here

- **Are they correlated?** (next slide)



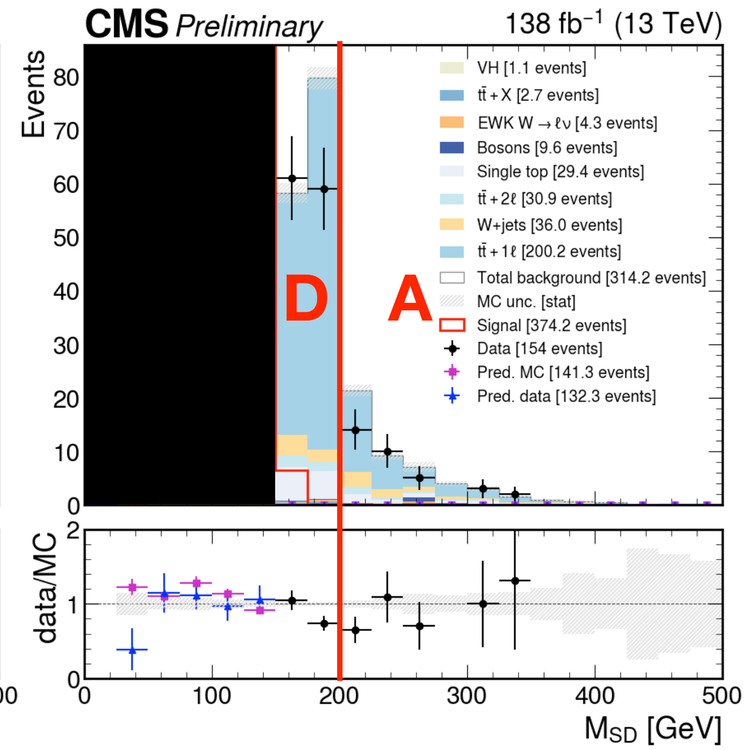
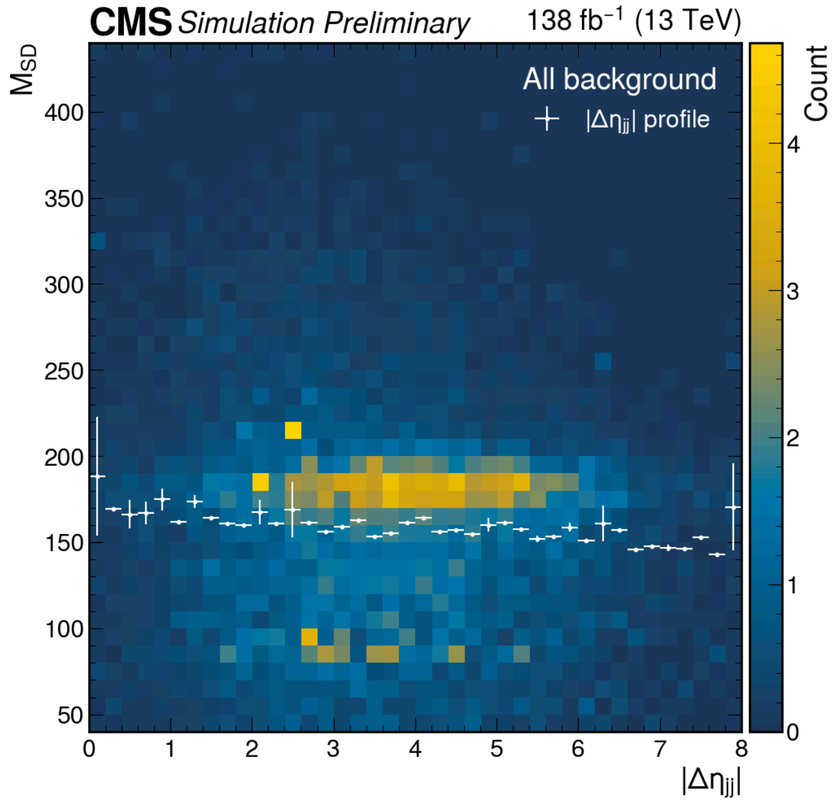
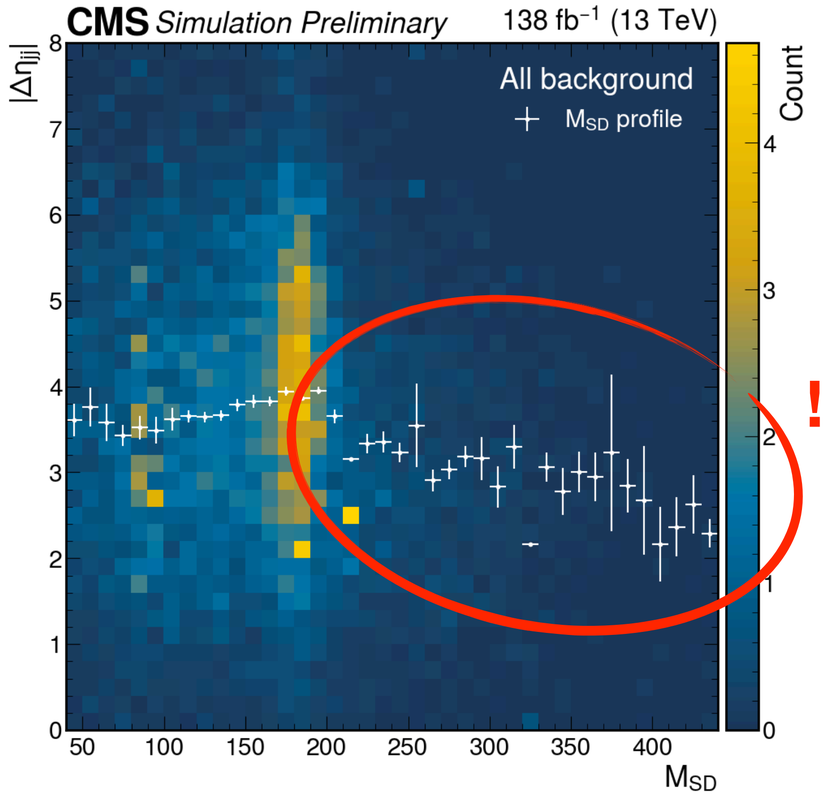


*Table errors = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego

2D Correlation Plots

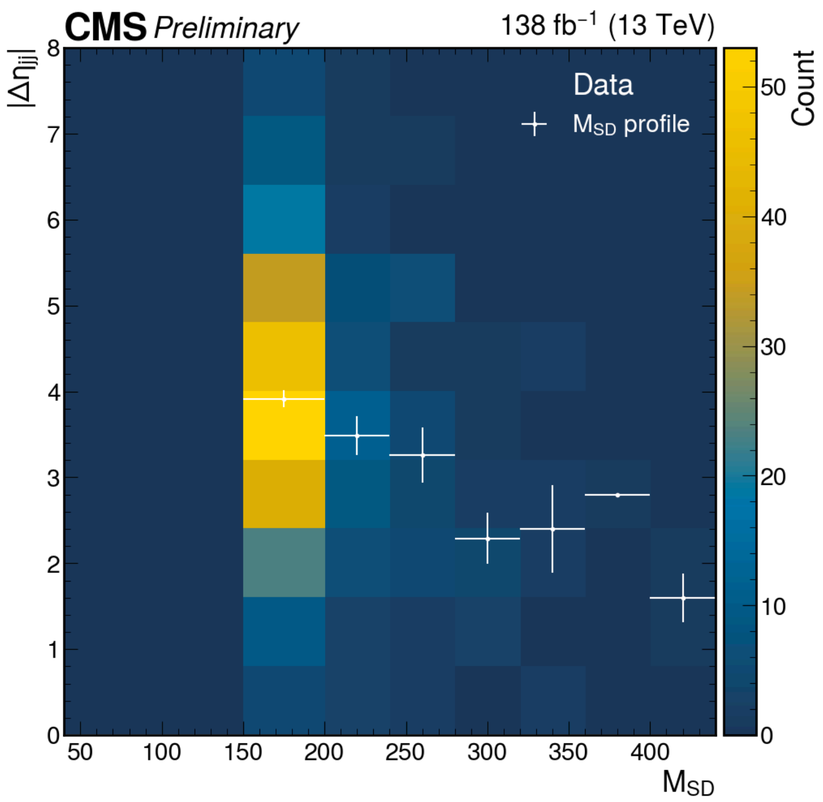


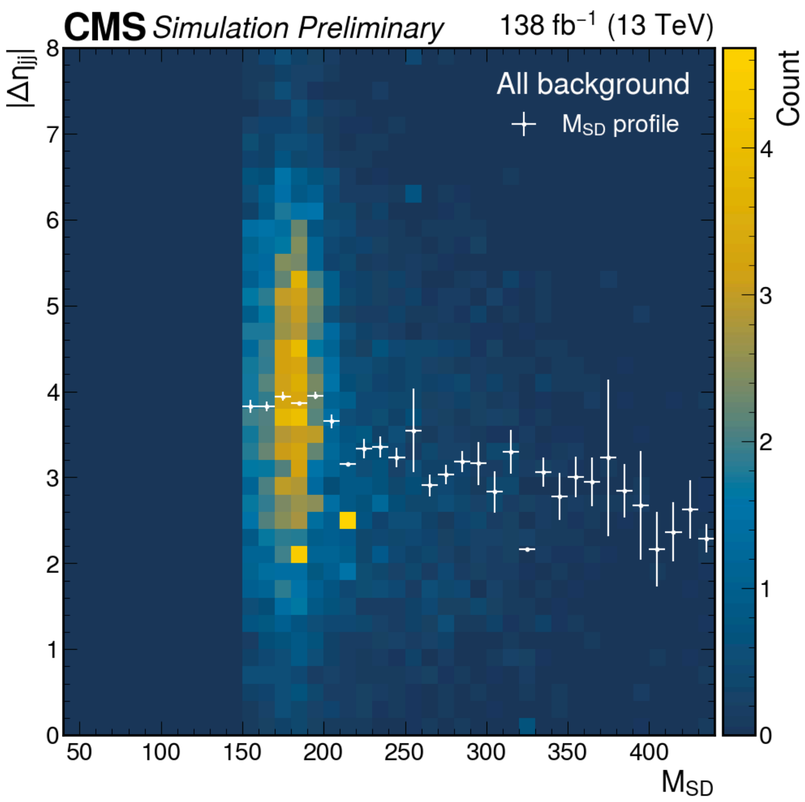


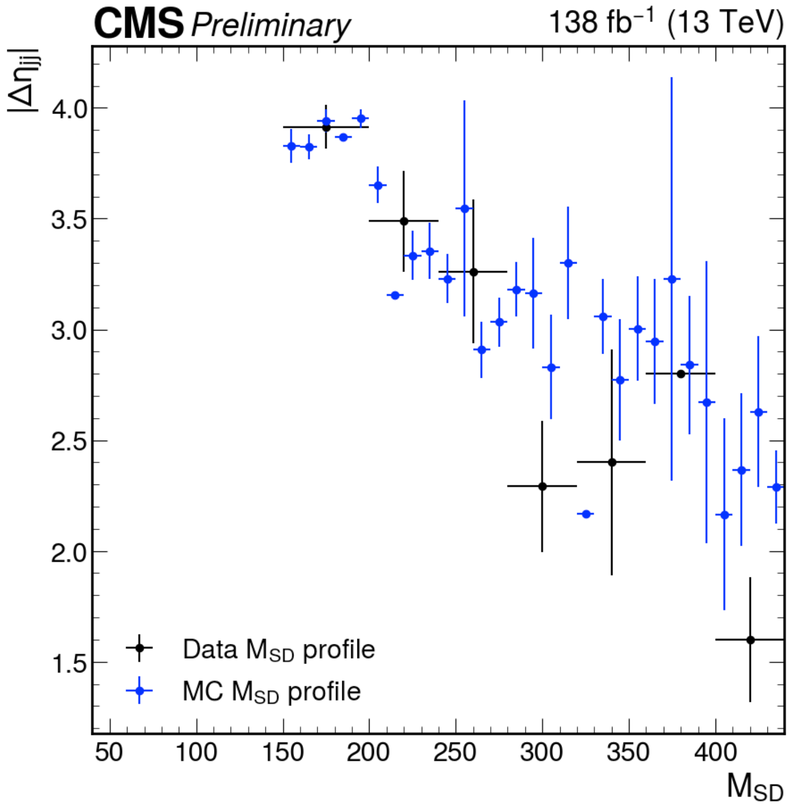
- Left plot looks OK, but correlation is clear in right plot
- **M_{SD} and $|\Delta\eta_{jj}|$ have some correlation (but "cancel out" for ABCD as defined)**

UC San Diego

2D Correlation Plots: Data vs. MC







- Trend in 1D profile is ~consistent between data and MC
- **Correlation is well-modeled in MC → a correction can be taken from MC (next slide)**

UC San Diego

Proposed Correction

Before: $D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}}$

After: $D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}} \times \left(\frac{D_{MC}/C_{MC}}{A_{MC}/B_{MC}} \right)$

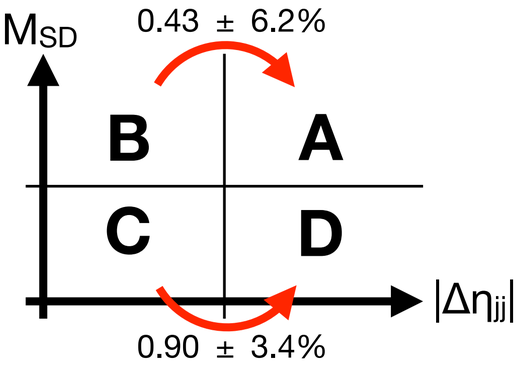
- We propose a **correction factor** taken from MC:
  - We know A/B ≠ D/C (hence our 11% systematic)
  - We compute a factor from MC that scales A/B to D/C (i.e. trivial closure in MC)
  - We apply this factor to A/B in data
- **Next: we test this procedure in the M_{SD} sideband**

With Correction: ABCD in M_{SD} Sideband

Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 AND M_{SD} ≥ 150 GeV

Cut	Region	Total Bkg.	Sig. (λ_{WZ} = -1)	Data
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} ≥ 200 GeV	A	41.8 ± 1.7	0.3 ± 1.7	28 ± 5.3
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND $M_{SD}\in$ ≥ 200 GeV	B	96.1 ± 4.4	0.0 ± 4.4	70 ± 8.4
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} < 200 GeV	C	145.9 ± 3.8	0.2 ± 3.8	131 ± 11.5
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} < 200 GeV ("SR")	D	131.2 ± 2.8	6.6 ± 2.8	**114** ± 10.7

*



- Good closure in data after applying transfer factor proposed on previous slide
- This confirms correlation is sufficiently well-modeled by MC
- We now apply this correction to the original ABCD

$$D^{pred}_{data} = C_{data} \times \frac{A_{data}}{B_{data}} \times \left(\frac{D_{MC}/C_{MC}}{A_{MC}/B_{MC}} \right)$$

$$= 131 \times \frac{28}{70} \times \left(\frac{0.899}{0.435} \right)$$

$$= 131 \times 0.400 \times \left(2.069 \right)$$

$$= \boxed{108.4}$$

*Table errors = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

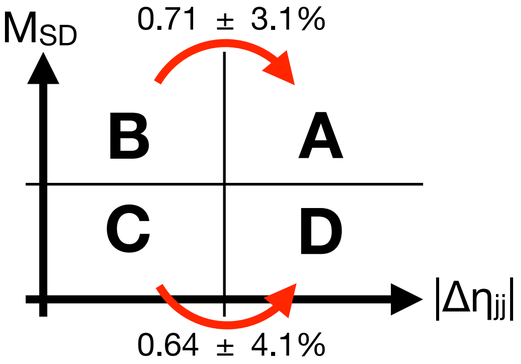
UC San Diego

With Correction: Original ABCD Method

Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Total Bkg.	Sig. (λ_{WZ} = -1)	Data
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} ≥ 150 GeV	A	173.0 ± 3.3	6.9 ± 3.3	142 ± 11.9
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	241.9 ± 5.8	0.3 ± 5.8	201 ± 14.2
$\lvert\Delta\eta_{jj}\rvert$ ≤ 4 AND M_{SD} < 150 GeV	C	181.1 ± 4.4	11.6 ± 4.4	170 ± 13.0
$\lvert\Delta\eta_{jj}\rvert$ > 4 AND M_{SD} < 150 GeV (SR)	D	116.4 ± 3.8	366.3 ± 3.8	–

*



- Only slightly different predicted background yield
  - Consistent within 1σ data statistical error
- Keep systematic uncertainties from before (backup)

$$D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}} \times \left(\frac{D_{MC}/C_{MC}}{A_{MC}/B_{MC}} \right)$$

$$= 170 \times \frac{142}{201} \times \left(\frac{0.643}{0.715} \right)$$

$$= 170 \times 0.706 \times \left(\cdot \quad 0.899 \right)$$

$$= \mathbf{108.0}$$

Pred. bkg. before: 120.1 ± 16.07 ± 15.30
stat. *syst.*

Pred. bkg. after: 108.0 ± 14.45 ± 13.76
stat. *syst.*

*Table errors = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego

Summary

- After producing the correlation plots requested by L2s, we see that our ABCD variables are correlated
  - However, this correlation “cancels out” and we get good closure in MC
  - Correlation is also well-modeled in data
- We propose a **correction factor** from MC:

$$D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}} \times \left(\frac{D_{MC}/C_{MC}}{A_{MC}/B_{MC}} \right)$$

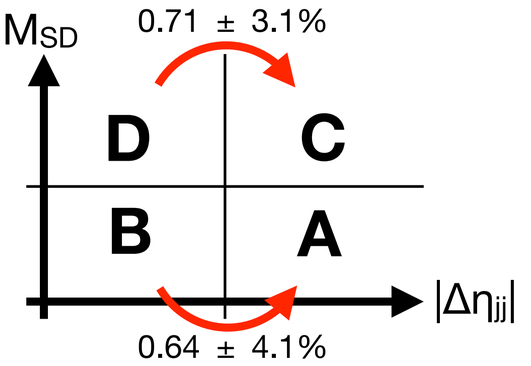
- We confirm that this is valid in M_{SD} sideband (i.e. now closes in data)
- We apply this correction factor to our original ABCD, get only a slightly different prediction, and keep systematics from before

Backup

ABCD Background Estimation

Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} ≥ 150 GeV	A	172.97	3.25	6.92	3.25	142	11.92
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	241.93	5.83	0.27	5.83	201	14.18
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} < 150 GeV	C	181.10	4.40	11.62	4.40	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	**116.41**	3.84	366.30	3.84	–	–



- Errors: 10% (syst.), 13% (stat.)

Over-predicted

$$D_{MC}^{pred} = \frac{A_{MC}}{B_{MC}} \times C_{MC} = \mathbf{129.48}$$

$$D_{data}^{pred} = \frac{A_{data}}{B_{data}} \times C_{data} = 120.10$$

Bkg comp. syst. (backup)

$$\epsilon_{syst} = \left| 1 - \frac{D_{MC}^{pred}}{D_{MC}} \right| = \left| 1 - \frac{129.5}{116.4} \right| = 11\% \oplus 6\% = 13\%$$

$$\epsilon_{stat} = \sqrt{\left(\frac{\sqrt{A_{data}}}{A_{data}}\right)^2 + \left(\frac{\sqrt{B_{data}}}{B_{data}}\right)^2 + \left(\frac{\sqrt{C_{data}}}{C_{data}}\right)^2}$$

$$= \sqrt{\frac{1}{A_{data}} + \frac{1}{B_{data}} + \frac{1}{C_{data}}} = 13\%$$

Predicted SR Yield: $120.1 \pm 16.07 \pm 15.30$

stat. *syst.*

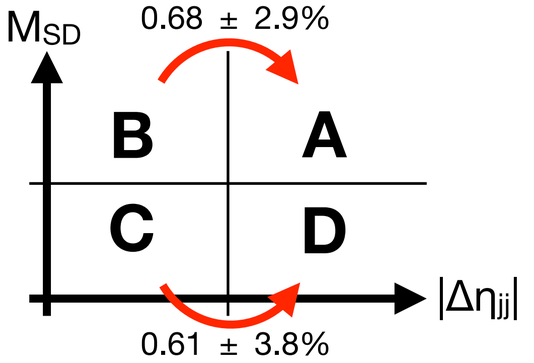
*err = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego

SR1 ABCD: W+jets Composition

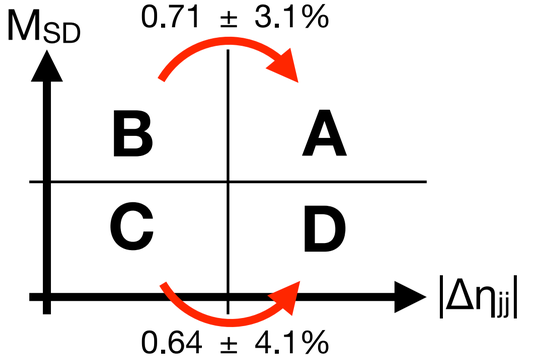
Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 (WJets x 2)

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} ≥ 150 GeV	A	184.26	3.48	6.92	0.40	142	11.92
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	272.50	5.98	0.27	0.08	201	14.18
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} < 150 GeV	C	223.95	4.72	11.62	0.52	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	137.64	4.42	366.30	2.92	—	—



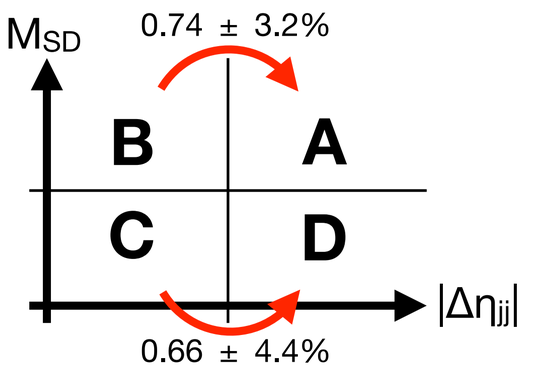
Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} ≥ 150 GeV	A	172.97	3.25	6.92	3.25	142	11.92
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	241.93	5.83	0.27	5.83	201	14.18
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} < 150 GeV	C	181.10	4.40	11.62	4.40	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	116.41	3.84	366.30	3.84	—	—



Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 (WJets x 0.5)

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} ≥ 150 GeV	A	167.32	3.19	6.92	0.40	142	11.92
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} ≥ 150 GeV	B	226.65	5.79	0.27	0.08	201	14.18
$\|\Delta\eta_{jj}\|$ ≤ 4 AND M_{SD} < 150 GeV	C	159.67	4.32	11.62	0.52	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	105.79	3.68	366.30	2.92	—	—



5.4% systematic

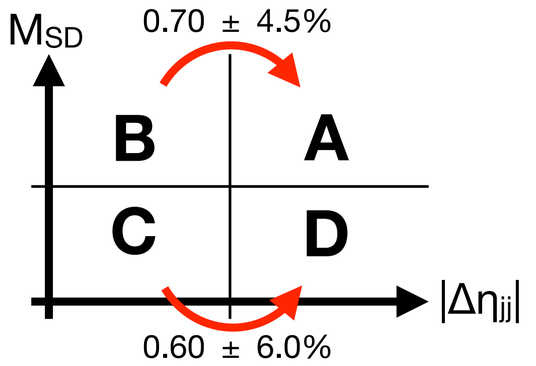
*err = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego

SR1 ABCD: Bosons Composition

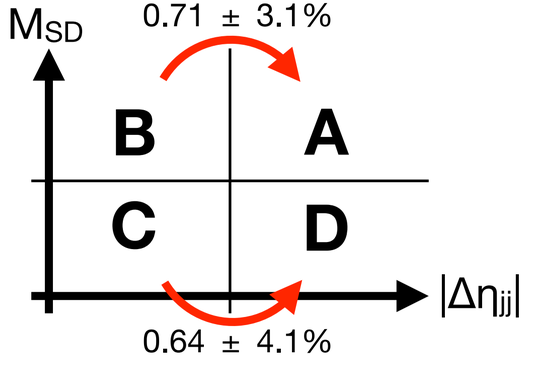
Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 (Bosons x 2)

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND $M_{SD} \geq$ 150 GeV	A	173.96	3.46	6.92	0.40	142	11.92
$\|\Delta\eta_{jj}\| \leq$ 4 AND $M_{SD} \geq$ 150 GeV	B	249.87	9.99	0.27	0.08	201	14.18
$\|\Delta\eta_{jj}\| \leq$ 4 AND M_{SD} < 150 GeV	C	202.63	7.17	11.62	0.52	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	122.39	5.97	366.30	2.92	—	—



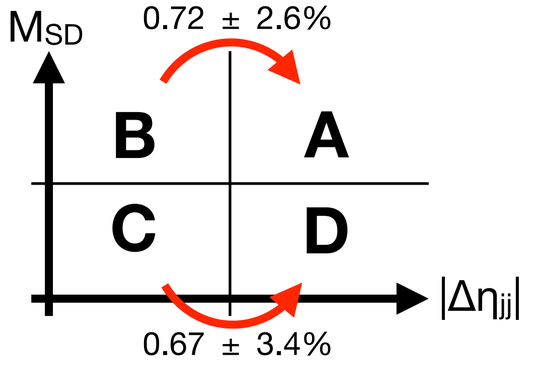
Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND $M_{SD} \geq$ 150 GeV	A	172.97	3.25	6.92	3.25	142	11.92
$\|\Delta\eta_{jj}\| \leq$ 4 AND $M_{SD} \geq$ 150 GeV	B	241.93	5.83	0.27	5.83	201	14.18
$\|\Delta\eta_{jj}\| \leq$ 4 AND M_{SD} < 150 GeV	C	181.10	4.40	11.62	4.40	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	116.41	3.84	366.30	3.84	—	—



Preselection AND M_{jj} > 600 GeV AND S_T > 900 GeV AND PNet Xbb > 0.9 (Bosons x 0.5)

Cut	Region	Bkg. (wgt)	Bkg. Err.*	Sig. (wgt)	Sig. Err.*	Data	Data Err.*
$\|\Delta\eta_{jj}\|$ > 4 AND $M_{SD} \geq$ 150 GeV	A	172.47	3.20	6.92	0.40	142	11.92
$\|\Delta\eta_{jj}\| \leq$ 4 AND $M_{SD} \geq$ 150 GeV	B	237.97	4.18	0.27	0.08	201	14.18
$\|\Delta\eta_{jj}\| \leq$ 4 AND M_{SD} < 150 GeV	C	170.33	3.38	11.62	0.52	170	13.04
$\|\Delta\eta_{jj}\|$ > 4 AND M_{SD} < 150 GeV (SR)	D	113.42	3.08	366.30	2.92	—	—



2.6% systematic

*err = $\sqrt{(\Sigma_i w_i^2)}$ for MC, $\sqrt{(\text{count})}$ for data

UC San Diego