



Andrea Castro  
- University of Bologna and INFN -



# Measurements of $t\bar{t}$ +jets, $t\bar{t}$ +bb and $t\bar{t}t\bar{t}$ production at CMS



*On behalf of the CMS collaboration*

ICHEP-2020

# Why $t\bar{t} + xx$ ?

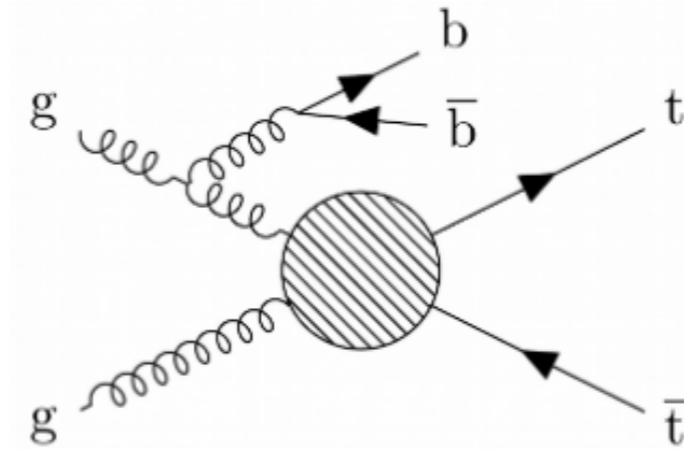
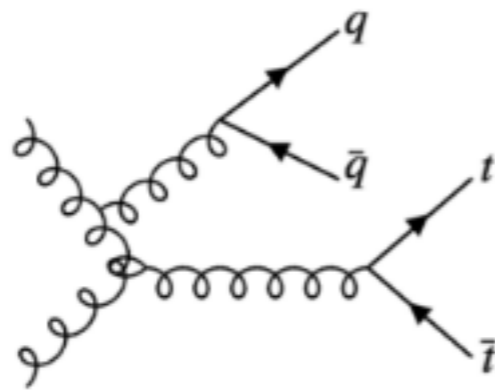
Top quark physics lies at the edge of the SM territory so it is important to map this area as fully as possible

$t\bar{t}$  production is abundant at the LHC and quite well covered (total and differential cross sections) so the focus shifts to  $t\bar{t} + xx$  production

$t\bar{t} + xx$  events are important as background for rare SM processes but might be also related to new physics BSM

$t\bar{t} + xx$  production can also be interpreted using EFT

# Why $ttjj$ , $ttbb$ ?



Important background for  $ttH$  and for BSM events

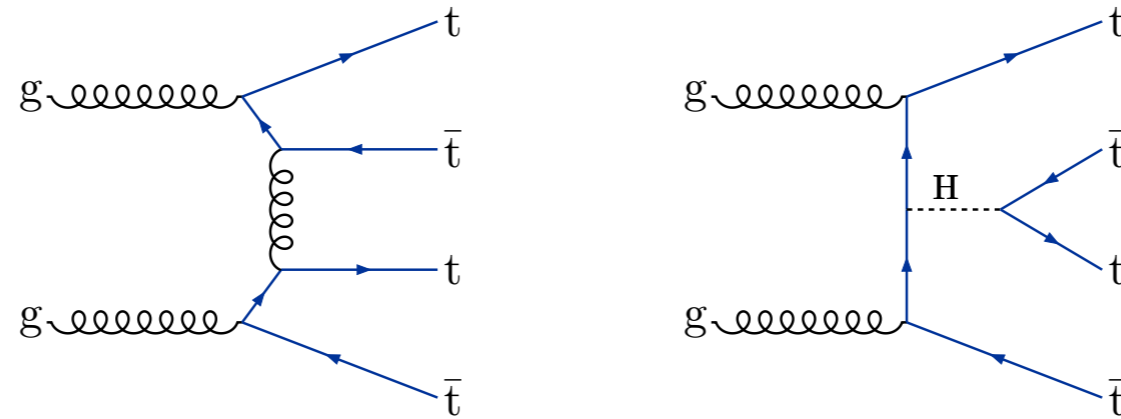
Precise measurements needed to improve  $tt$ +jets MC simulation

NLO calculations are affected by large uncertainties associated to  $\mu_R/\mu_F$  scales; two different scales:  $m_t$  vs  $p_T$

Different energy scales for associated b-jets and jets from the  $tt$  system

Different modeling approaches: NLO  $tt$ +jets (5FS, massless b quarks), NLO  $ttbb$  (4FS, massive b quarks)

# Why $t\bar{t}t\bar{t}$ ?



This production has not been "observed" yet, but is expected in the SM with a NLO cross section of  $\approx 12$  fb

Allows to constrain t-H Yukawa coupling (but less precisely than  $t\bar{t}H$ )

This channel is sensitive to new physics (BSM) e.g. two-Higgs-doublet models, SUSY, etc.

# Latest results on ttjj, ttbb, tttt production cross sections

We discuss here the most recent ones at 13 TeV:

- ttjj, ttbb (single lepton, dilepton),  $36 \text{ fb}^{-1}$
- ttbb (all-had),  $36 \text{ fb}^{-1}$
- tttt (single lepton, opposite-sign dilepton),  $36 \text{ fb}^{-1}$
- tttt (same-sign dilepton, trilepton),  $137 \text{ fb}^{-1}$



# ttjj + ttbb (lepton + dilepton) /1

2016 data (35.9 fb<sup>-1</sup> at 13 TeV)

Submitted to JHEP

## b-tagging discriminant for additional jets

### Selections

DL:

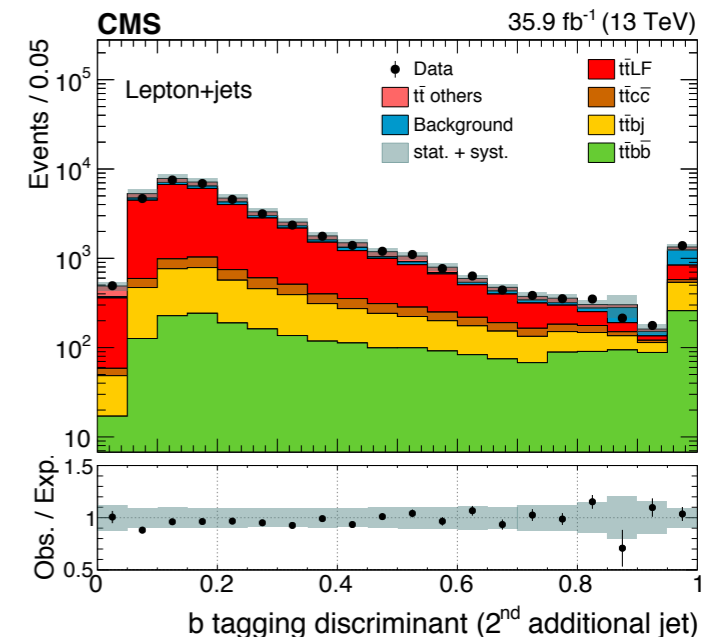
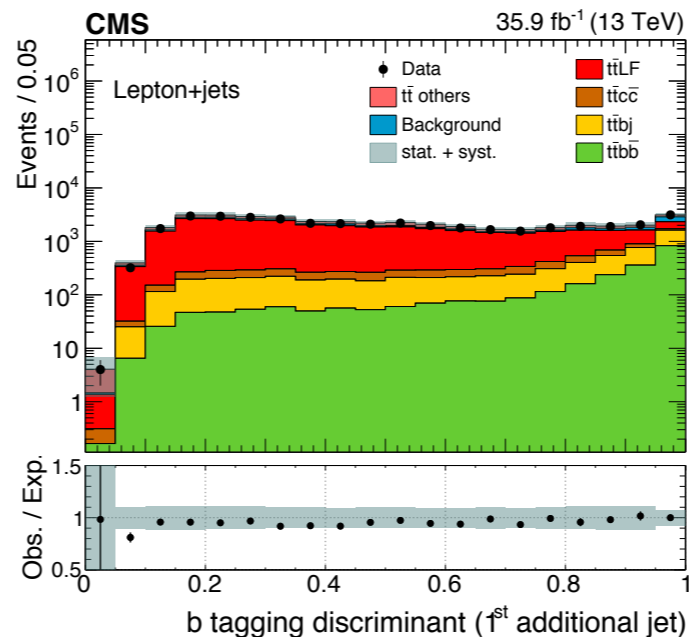
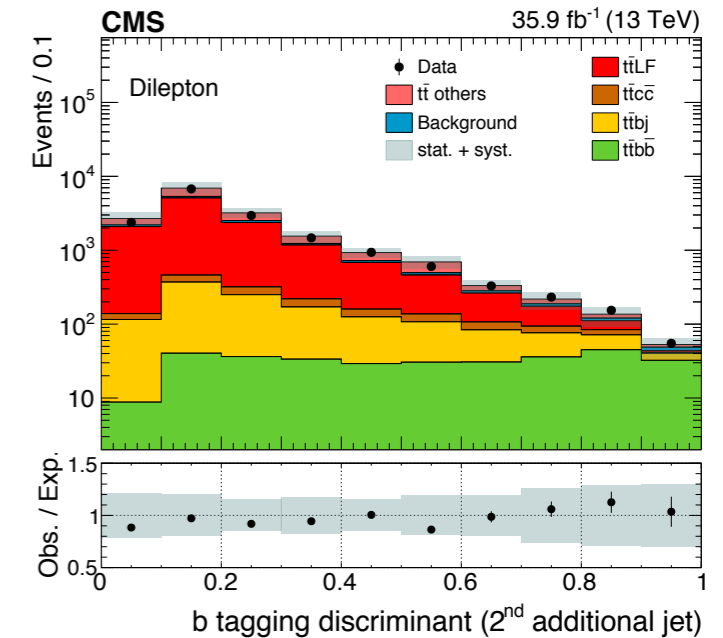
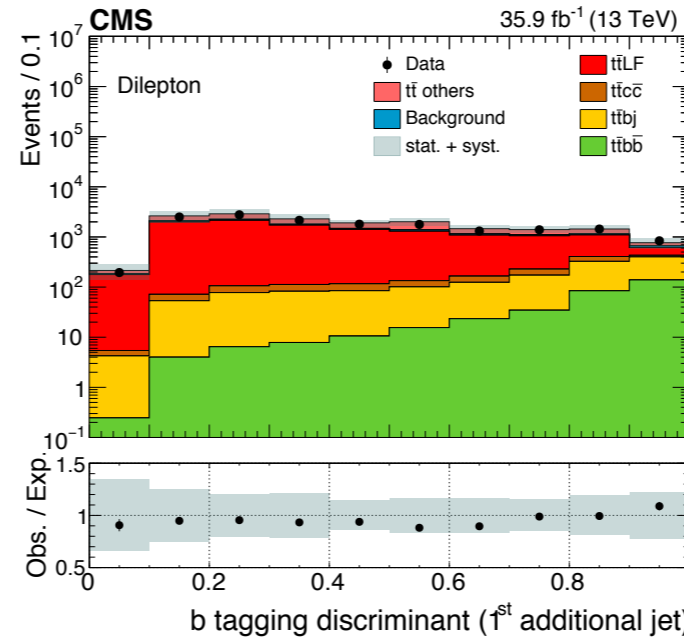
- 2 e/μ + ≥4 jets (p<sub>T</sub>>30 GeV)
- out of which ≥2 b-jets

SL:

- 1 e(μ) + ≥6 jets (p<sub>T</sub>>20 GeV)
- incl. ≥2 b-jets (tighter)

### Origin of additional jets:

- DL: 3rd and 4th b-jet based on tagging discriminants
- SL: more possibilities, so jet assignment based on a tt kinematic fit; the first 2 jets not included in the fit are chosen





# ttjj + ttbb (lepton + dilepton) /2

## Phase space:

- visible (VPS), number of particle-level leptons/jets matching the selection
- full (FPS), all decays, with at least two jj/bb additional jets

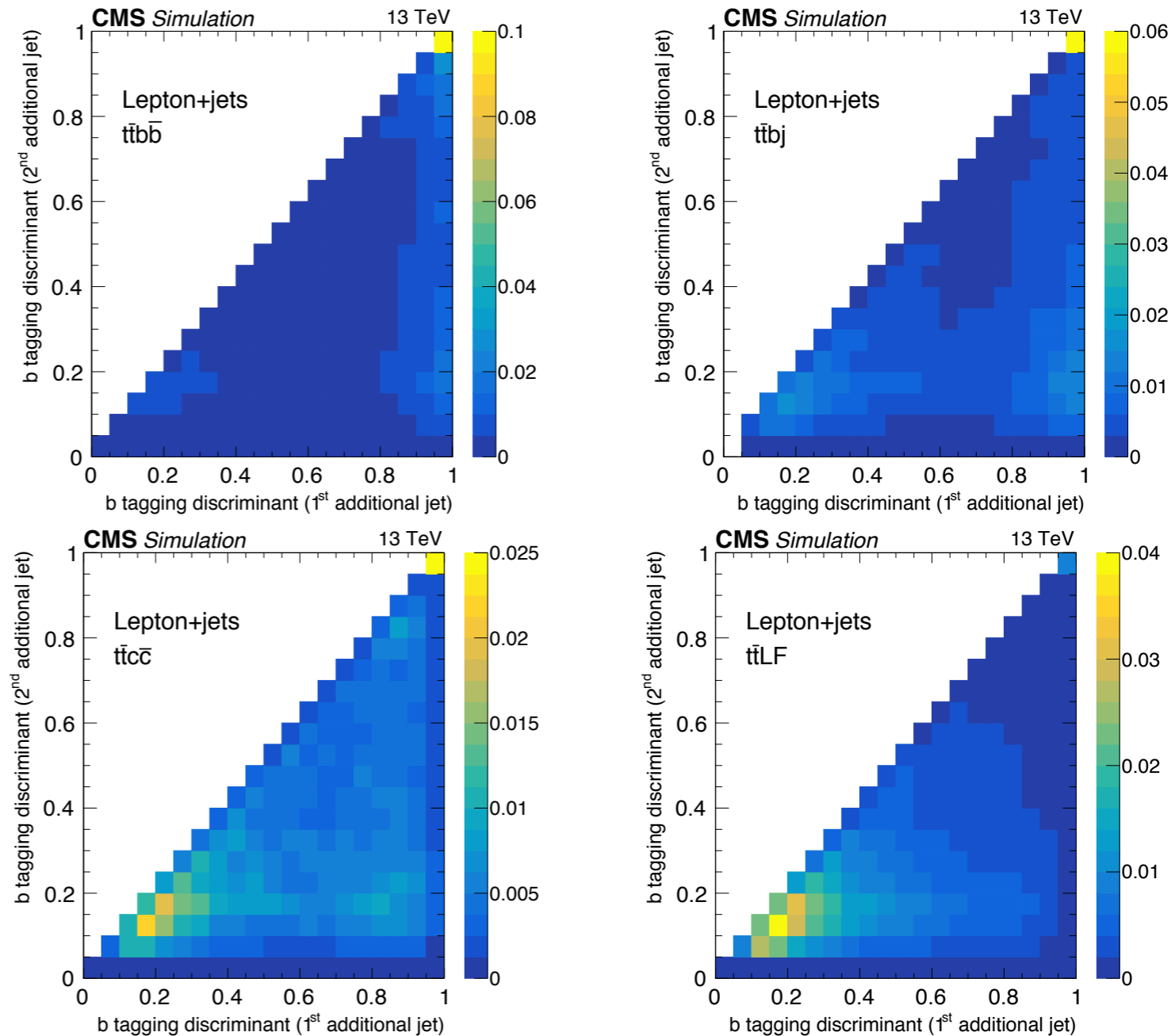
## Max likelihood binned fit in VPS to the 2D b-tagging discriminants:

- $\sigma(ttjj)$
- $R = \sigma(ttbb)/\sigma(ttjj)$

independently for the two channels (because of different phase space and backgrounds)

Then extrapolated to FPS using acceptances from MC

## 2D b-tagging discriminants for additional jets





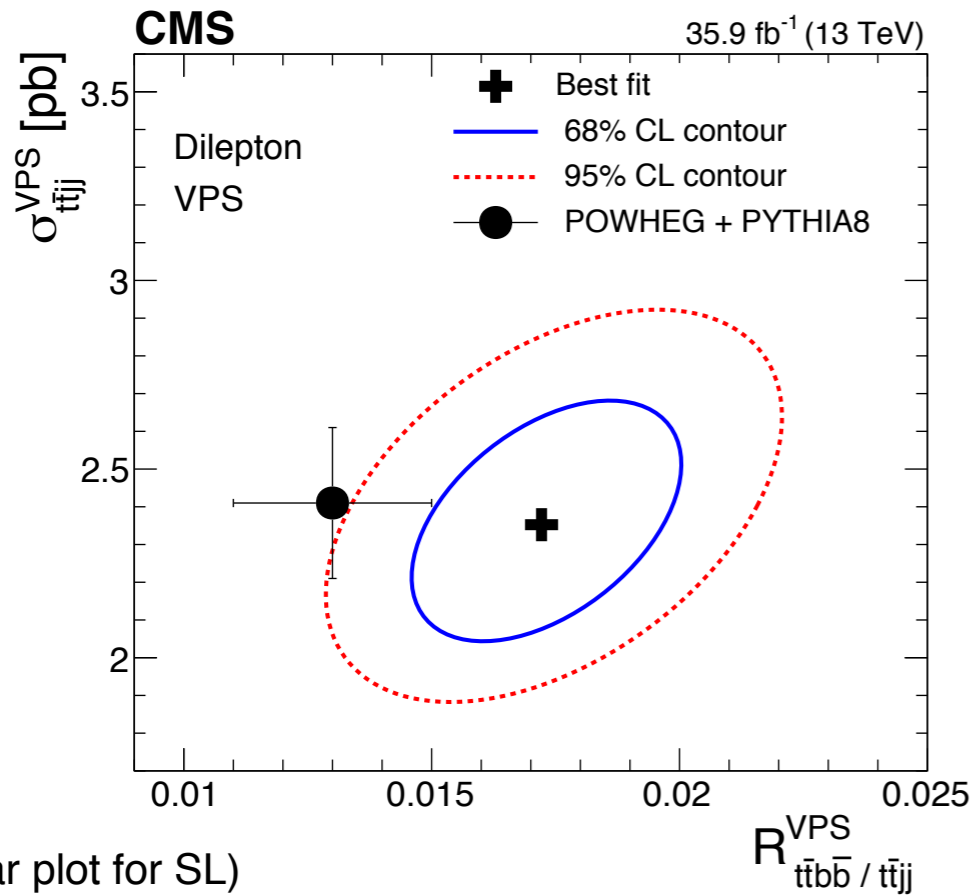
# ttjj + ttbb (lepton + dilepton) /3

## Cross sections / ratio (VPS):

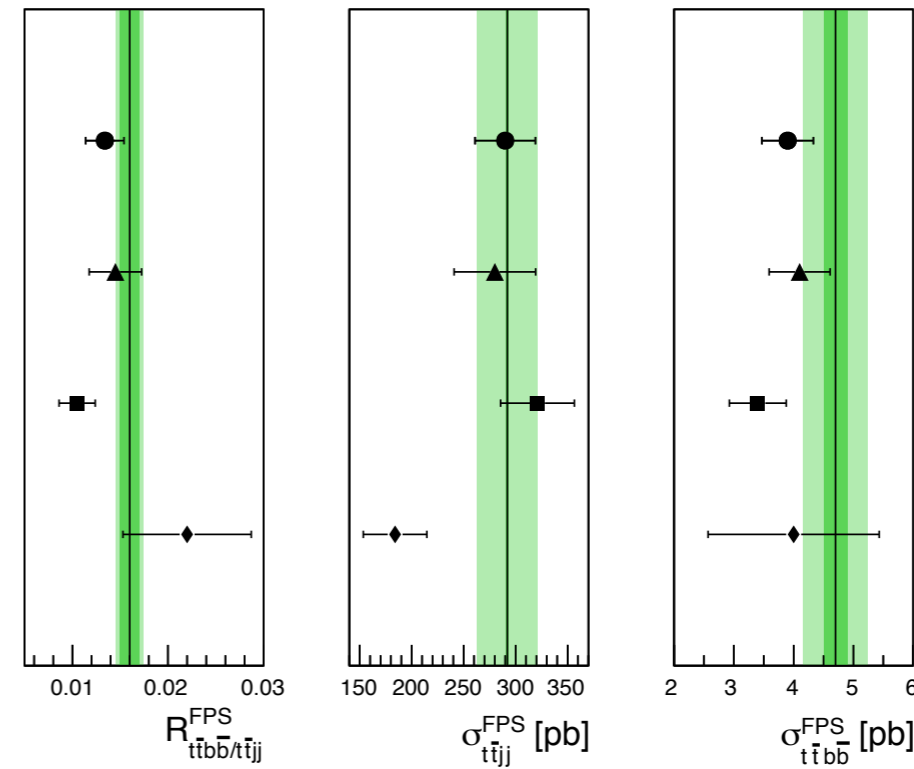
- $\sigma_{\text{obs}}(\text{ttjj}) \approx \sigma_{\text{SM}}(\text{ttjj})$
- $R_{\text{obs}}^{\text{DL}} \approx 1.3 \cdot R_{\text{SM}}$  with  $3\sigma$  significance ( $\approx 1.8$  in a previous DL with 1/15 of data,  $2\sigma$  sign.)

## Cross sections / ratio (FPS):

- $\sigma_{\text{obs}}(\text{ttjj})$  and  $R_{\text{obs}}$  consistent with different MCs (but worse for POWHEG +HERWIG)



(similar plot for SL)



(similar plot for DL)

## Main systematics on R or $\sigma(\text{ttjj})$ :

- light-flavor btag
- JES
- FSR



# ttbb (all-had) /1

2016 data (35.9 fb<sup>-1</sup> at 13 TeV)

[PLB 803 \(2020\) 135285](#)

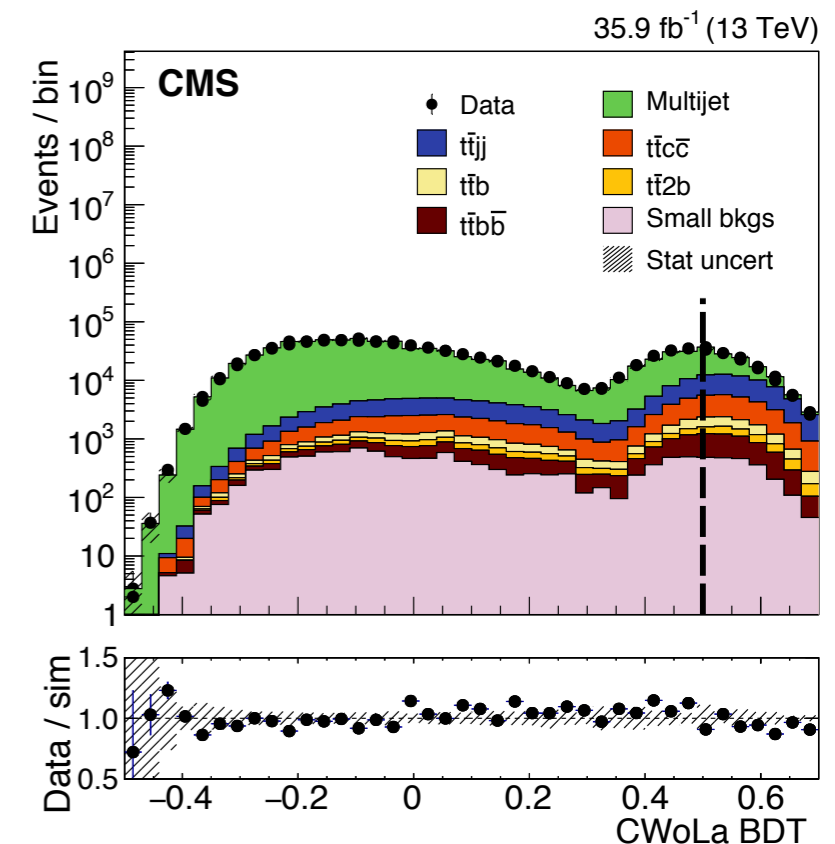
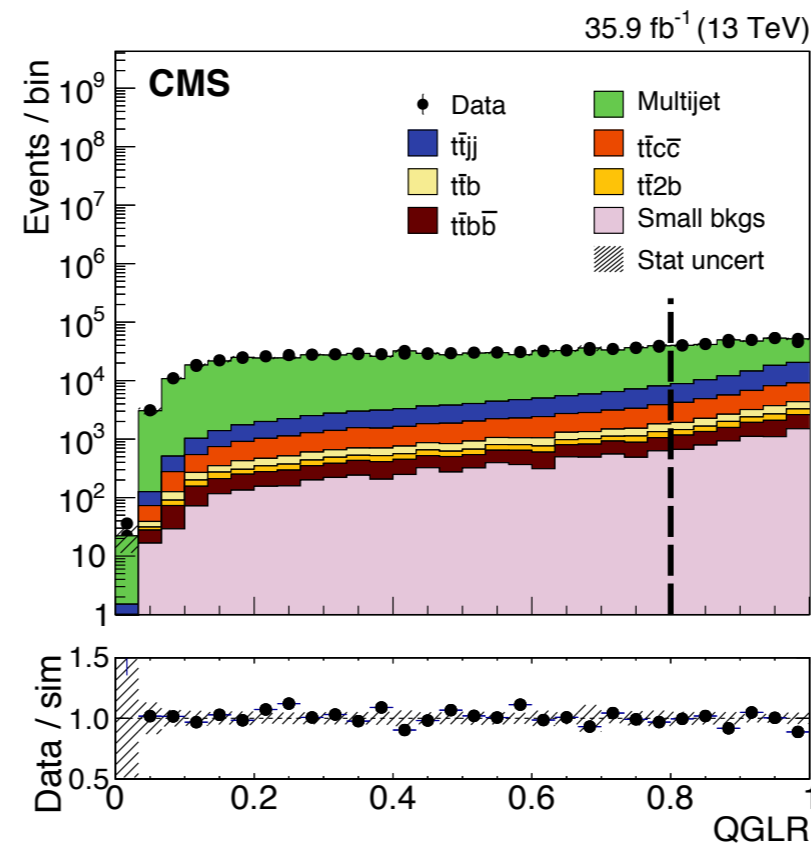
## Selection:

- $\geq 8$  jets
- out of which  $\geq 4$  b-jets
- (including 2 "additional" b-jets for the parton-based phase space)
- $QGLR > 0.8, CWoLa > 0.5$

## Main backgrounds:

- multijet (from data CRs)
- ttjj, ttcc

## MVA discriminators



## 3 MVA discriminators (to reduce bckgd and to identify additional jets):

1. quark-gluon likelihood ( $QGLR$ , based on jet substructure) favoring  $N_q=4$  vs  $N_g=4$
2. best jet permutation selected by BDT (based on jj and jjj inv. masses, correct at 60%)
3. weakly supervised BDT ( $CWoLa$ , based on data from signal/control regions) to reduce the multijet background



# ttbb (all-had) /2

2DCSV discriminators unrolled  
to one dimension and increasing S/B

## Two fiducial phase space definitions:

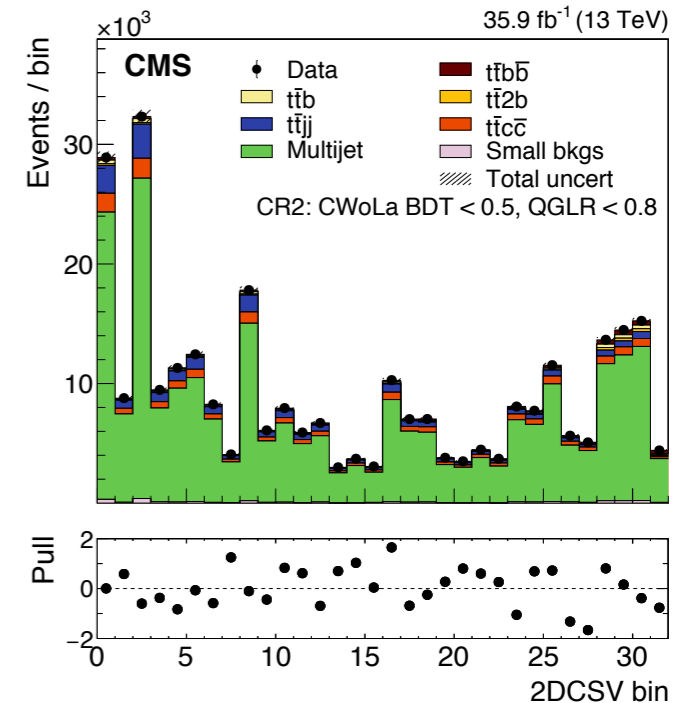
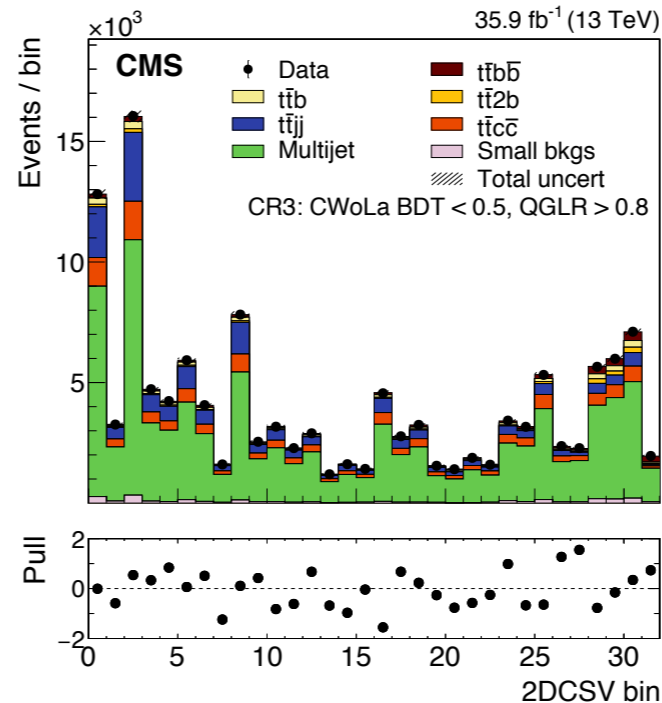
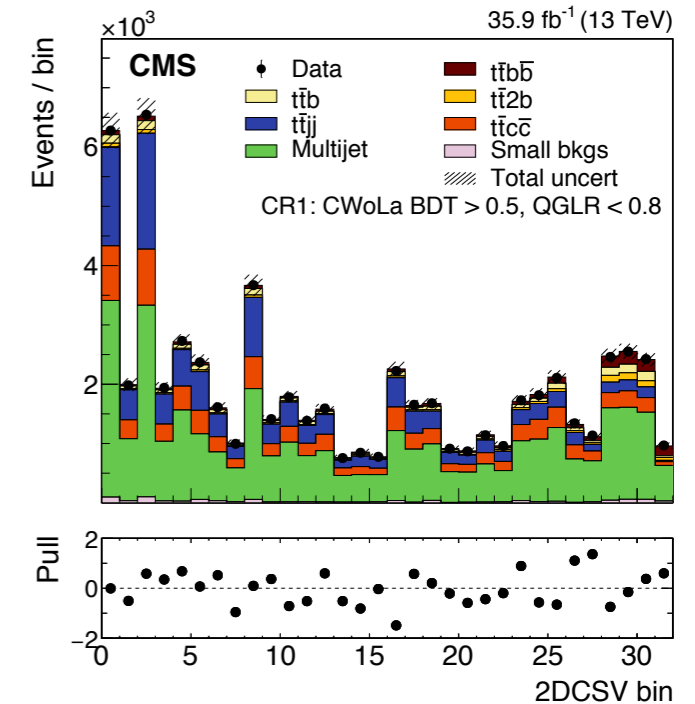
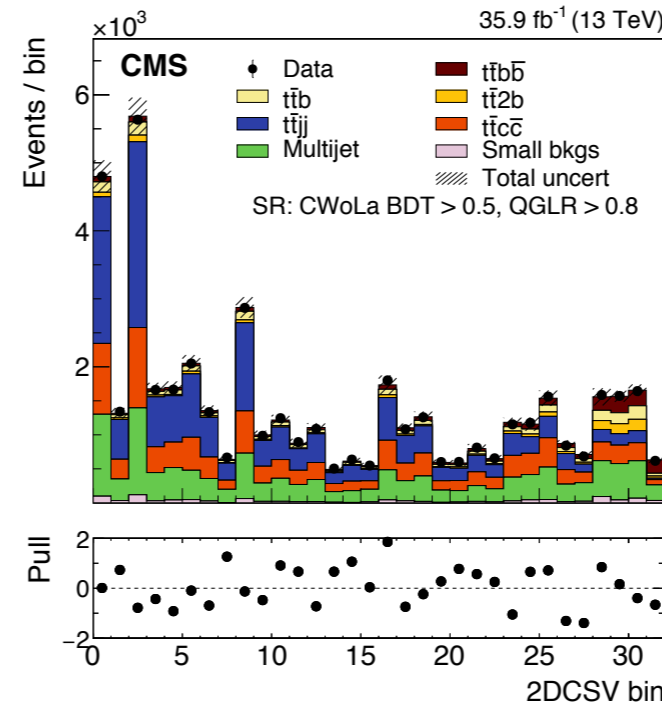
- parton-level after radiation (parton-based, PB)
- stable particles after hadronization (parton-independent, PI)

Total phase space is from PB after acceptance correction

Particle-level jets obtained clustering (anti- $k_T$ ) stable final-state particles

## 2DCSV discriminator:

- from largest and next b-tagging discriminator scores for "additional" b-jets (according to "permutation BDT")
- 4 analysis regions (SR, 3CRs)





# ttbb (all-had) /3

## Cross sections:

- data are fitted with a profiled ML technique to the 2DCSV in the 4 data regions

## Cross sections:

- predictions underestimate measurements by a factor 1.5-2.4 (1-2 st. dev.)
- consistent with previous results

	FPS PI (pb)	FPS PB (pb)	TPS (pb)
Measurement	$1.6 \pm 0.1^{+0.5}_{-0.4}$	$1.6 \pm 0.1^{+0.5}_{-0.4}$	$5.5 \pm 0.3^{+1.6}_{-1.3}$
POWHEG (tt)	$1.1 \pm 0.2$	$1.0 \pm 0.2$	$3.5 \pm 0.6$
POWHEG (tt) + HERWIG++	$0.8 \pm 0.2$	$0.8 \pm 0.2$	$3.0 \pm 0.5$
MADGRAPH5_aMC@NLO (4FS ttbb)	$0.8 \pm 0.2$	$0.8 \pm 0.2$	$2.3 \pm 0.7$
MADGRAPH5_aMC@NLO (5FS tt+jets, FxFx)	$1.0 \pm 0.1$	$1.0 \pm 0.1$	$3.6 \pm 0.3$

$$\sigma_{ttbb} = 5.5 \pm 0.3(\text{stat}) + 1.6 - 1.3(\text{syst}) \text{ pb}$$

## Main systematics on cross sections:

- simulated sample size ~ 15%
- q-g likelihood ~ 13%
- b-tagging ~ 10%

CMS

ttbb all-jet

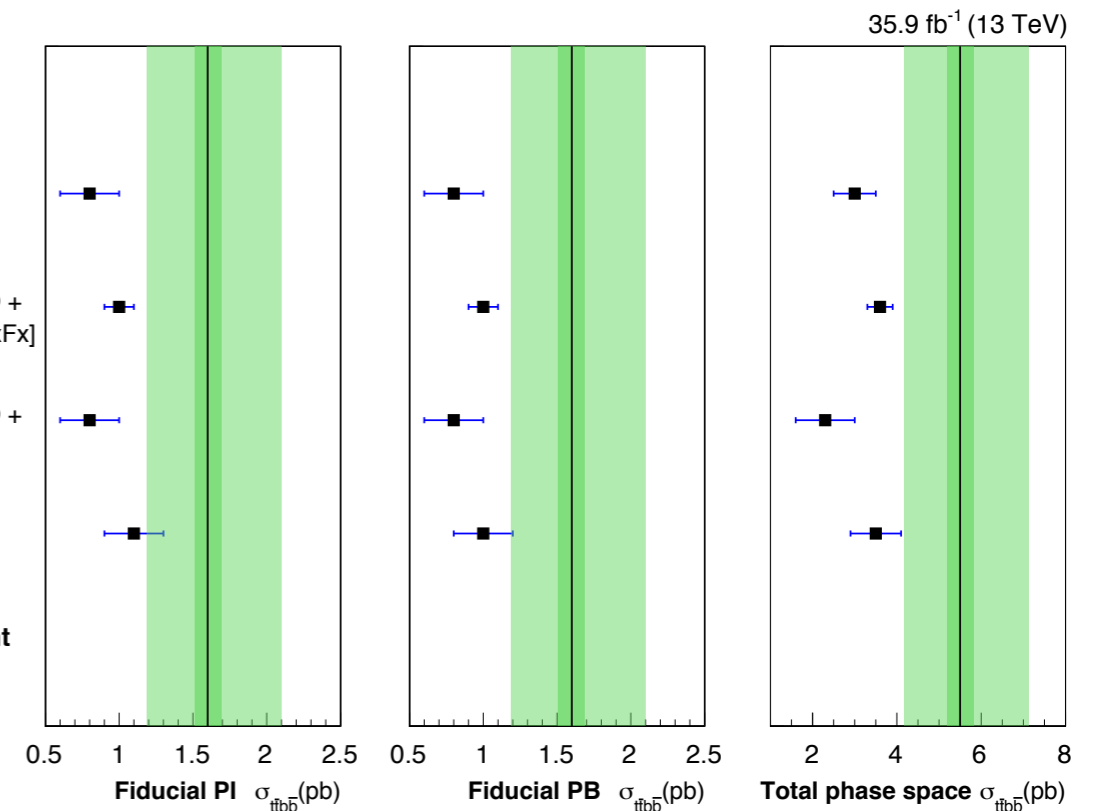
tt+jets:  
POWHEG +  
HERWIG++

tt+jets:  
MG5\_aMC@NLO +  
PYTHIA8 5FS [FxFx]

ttbb:  
MG5\_aMC@NLO +  
PYTHIA8 4FS

tt+jets:  
POWHEG +  
PYTHIA8

Measurement  
Total unc  
Stat unc





# $t\bar{t}t\bar{t}$ (single lepton + opposite-sign dilepton) /1

2016 data ( $35.8 \text{ fb}^{-1}$  at 13 TeV)

[JHEP 11 \(2019\) 082](#)

## Examples of OBS vs EXP distributions

### Selections

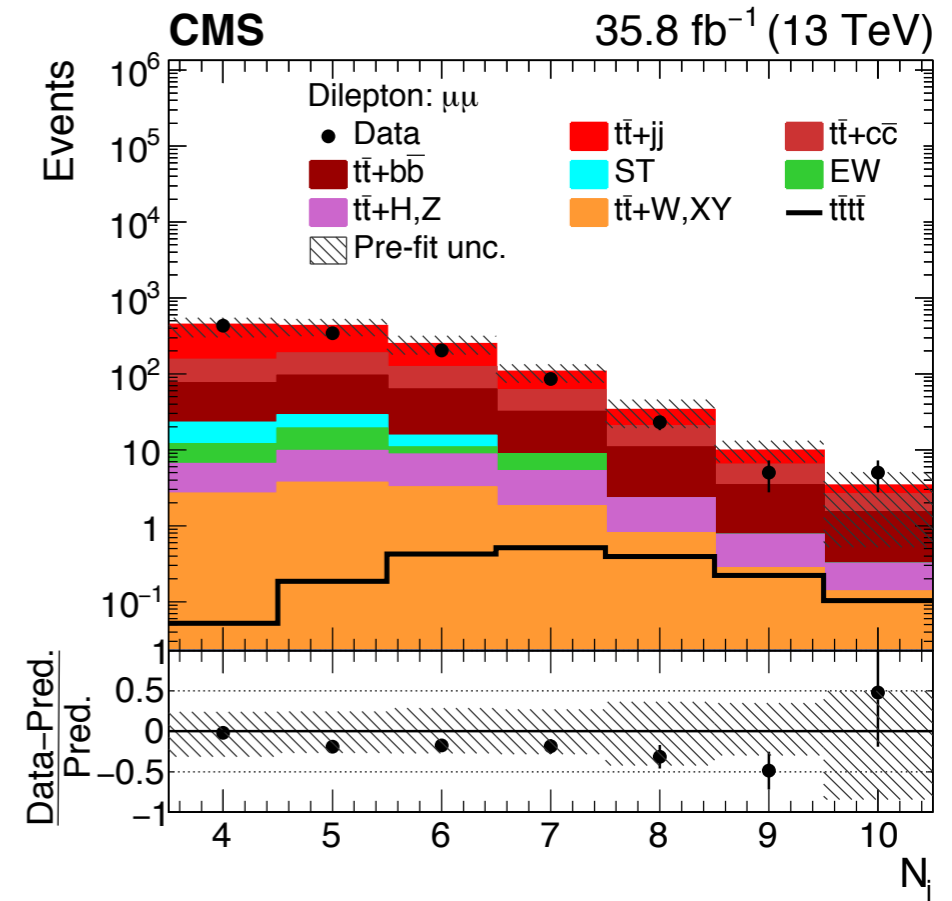
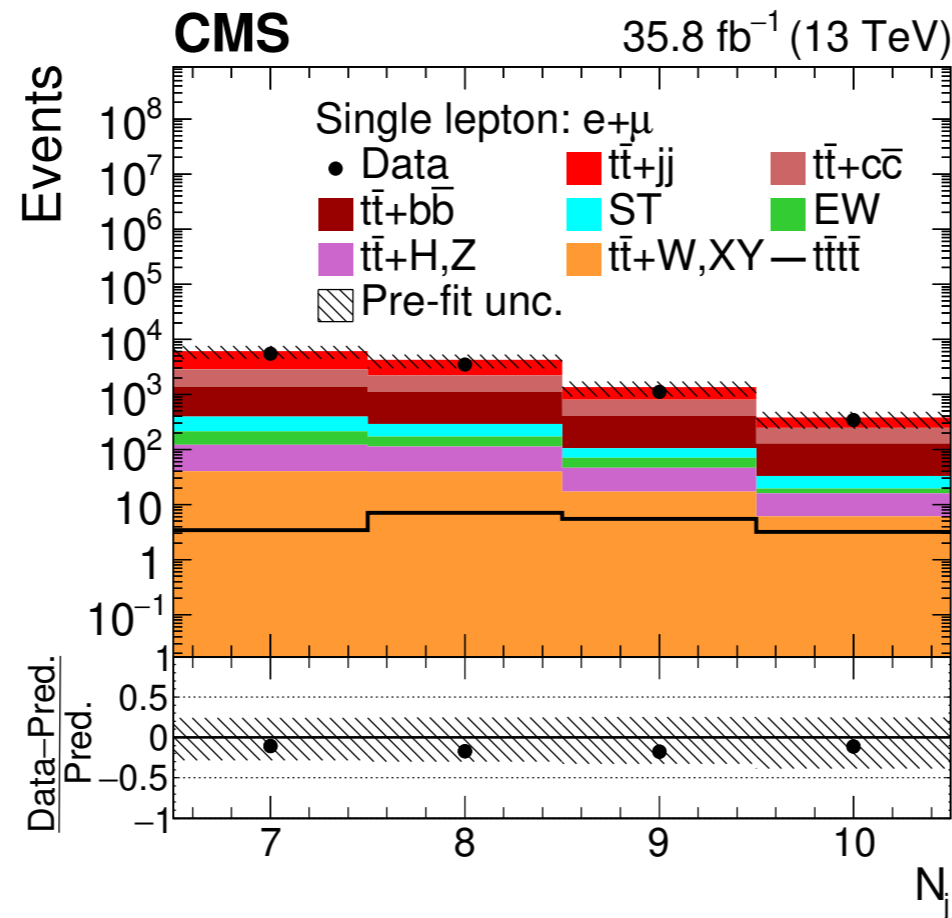
SL:

- 1  $e(\mu)$  +  $\geq 8(7)$  jets
- incl.  $\geq 2$  b-jets
- $H_T > 500 \text{ GeV}$

DL:

- 2  $e/\mu$  +  $\geq 4$  jets
- incl.  $\geq 2$  b-jets
- $H_T > 500 \text{ GeV}$

SL+DL  $\approx 40\%$  of  $t\bar{t}t\bar{t}$



- Main backgrounds:
- $t\bar{t}j\bar{j}$ ,  $t\bar{t}c\bar{c}$ ,  $t\bar{t}b\bar{b}$

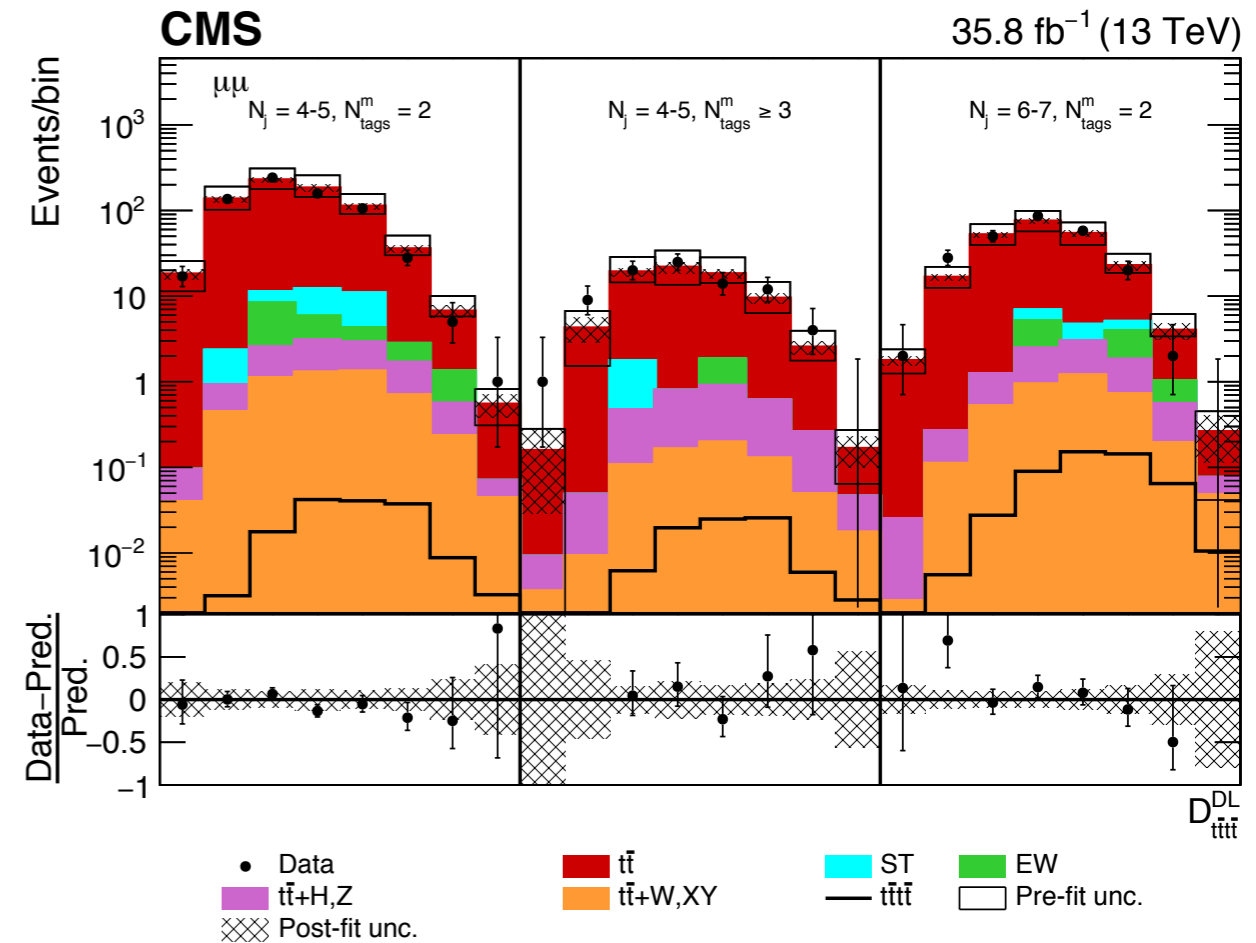
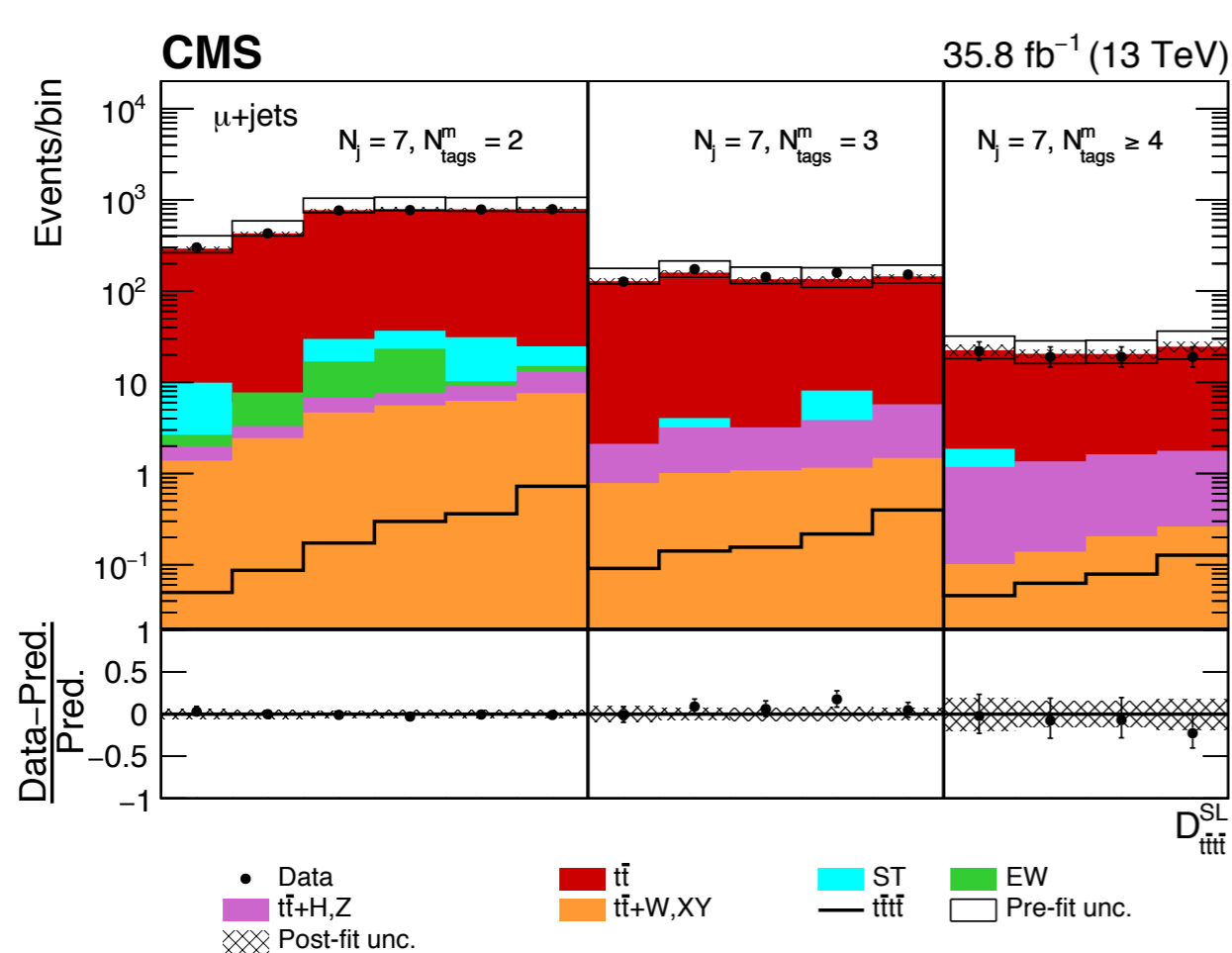


# $t\bar{t}t\bar{t}$ (single lepton + opposite-sign dilepton) /2

BDTs ( $D_{t\bar{t}t\bar{t}}^{SL}$ ,  $D_{t\bar{t}t\bar{t}}^{DL}$ ):

- used to identify the top quarks and to improve the S/B (one for each channel)
- to increase sensitivity, events are categorized based on  $N_j$  and  $N_b$

Two examples of OBS vs post-fit yields



No significant deviation is observed for all D distributions



# tttt (single lepton + opposite-sign dilepton) /3

Simultaneous binned ML template fit to extract:

- the signal strength  $\mu = \sigma_{tttt}^{obs}/\sigma_{tttt}^{SM}$
- the 95% CL upper limits on cross section

All of post-fit nuisances found consistent with initial values

Channel	Best fit $\mu$	Best fit $\sigma_{tt\bar{t}\bar{t}}$ (fb)	Exp. significance s.d.	Obs. significance s.d.
Single-lepton	$1.6^{+4.6}_{-1.6}$	$15^{+42}_{-15}$	0.21	0.36
Dilepton	$0.0^{+2.7}$	$0^{+25}$	0.36	0.0
Combined (this analysis)	$0.0^{+2.2}$	$0^{+20}$	0.40	0.0

no candidates



*For EFT interpretation, see talk by Samuel May (yesterday)*

Channel	Expected limit, $\mu$	Observed limit, $\mu$	Expected limit (fb)	Observed limit (fb)
Single-lepton	$9.4^{+4.4}_{-2.9}$	10.6	$86^{+40}_{-26}$	97
Dilepton	$7.3^{+4.5}_{-2.5}$	6.9	$67^{+41}_{-23}$	64
Combined (this analysis)	$5.7^{+2.9}_{-1.8}$	5.2	$52^{+26}_{-17}$	48

$\sigma_{tttt} < 48 \text{ fb at } 95\% \text{ CL}$

Systematics:

- many are driven by th. uncertainties e.g. tt+HF



# tttt (same-sign dilepton and trilepton) /1

2016-2018 data (137 fb<sup>-1</sup> at 13 TeV)

EPJC 80 (2020) 75

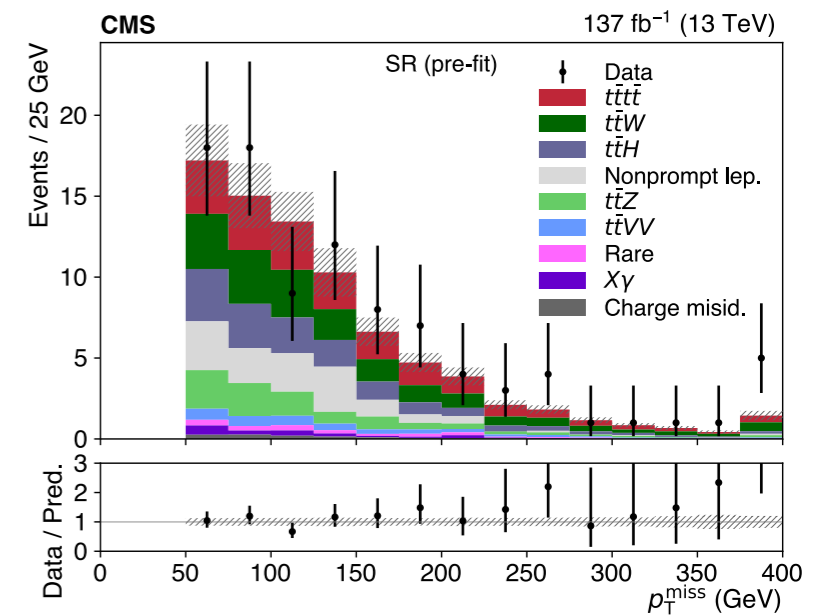
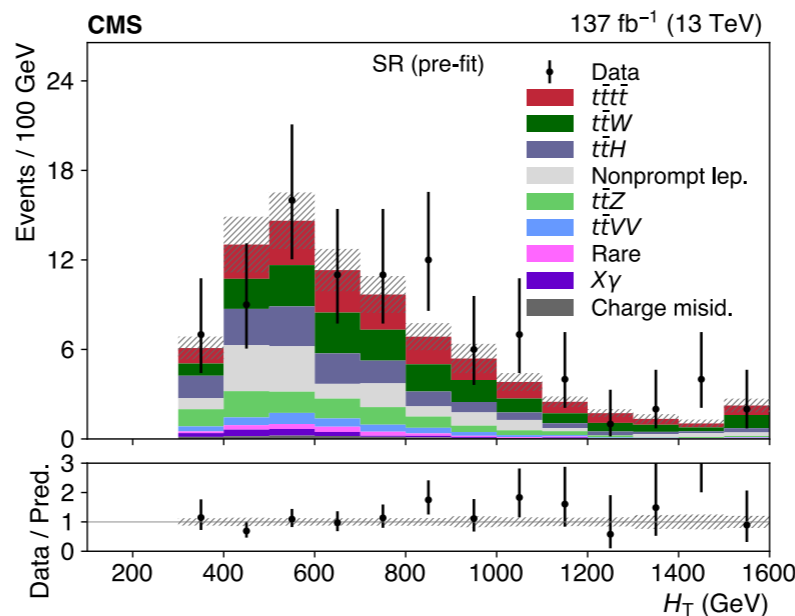
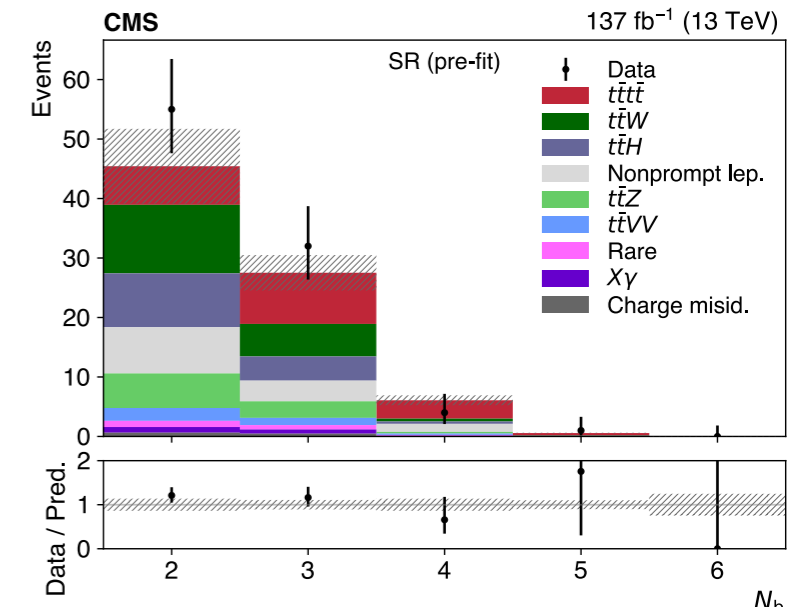
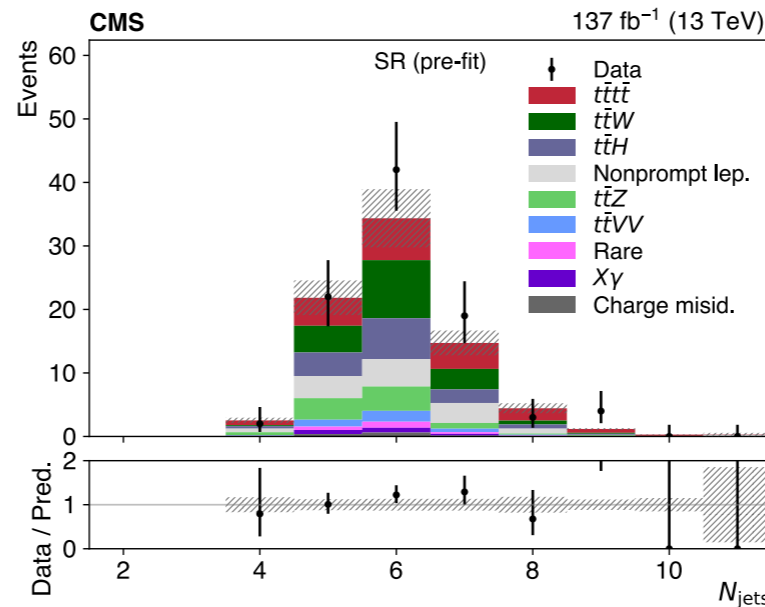
$N_j$ ,  $N_b$ ,  $H_T$ ,  $p_T^{\text{miss}}$  distributions

## Baseline selection:

- 2 e/ $\mu$  with same charge, or 3 e/ $\mu$  ( $p_T > 20$  GeV)
  - $\geq 2$  jets ( $p_T > 40$ ) +  $\geq 2$  b-jets ( $p_T > 25$ )
  - $p_T^{\text{miss}} > 50$  GeV
  - $H_T > 300$  GeV
- $\approx 12\%$  of tttt

## Main backgrounds:

- ttW, ttZ, ttH (the simulations are corrected for OBS/MC discrepancies)
- nonprompt lepton background is data-based (from isolation control regions)





# $tttt$ (same-sign dilepton and trilepton) /2

## Two analyses:

- cut-based ( $N_j$ ,  $N_b$ ,  $N_{e/\mu}$ : 14 SRs, 2 CRs)
- BDT approach ( $N_j$ ,  $N_b$ ,  $N_{e/\mu}$ ,  $p_T^{\text{miss}}$ ,  $H_T$  + 14 other vars: 17 SRs, 1 CR)  
 $\Rightarrow$  more sensitive/precise

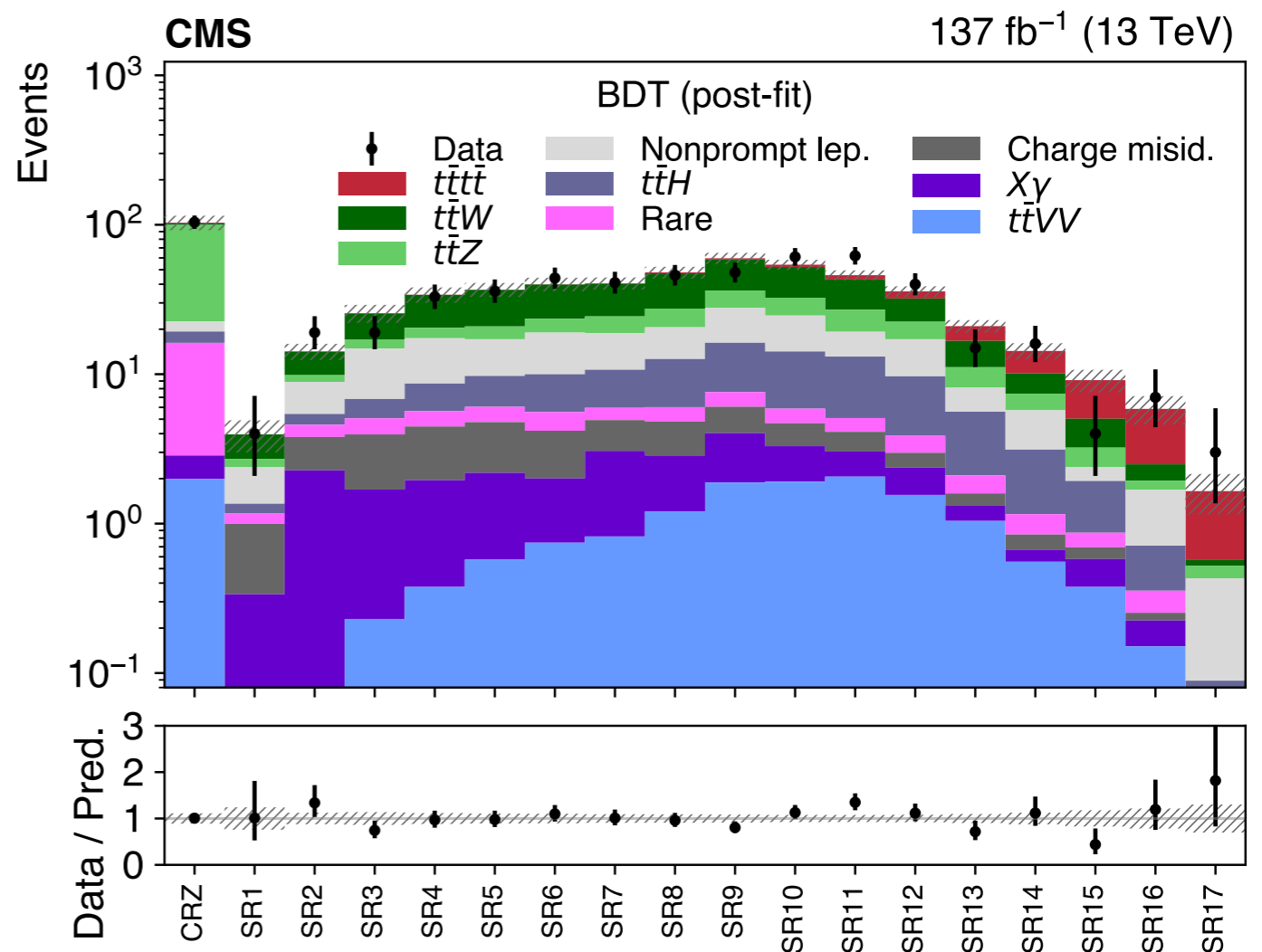
Observed and predicted yields enter a profiled ML fit to extract the cross section

Both analyses give consistent results but BDT analysis used for interpretations

## Main systematics on cross section:

- $\sigma(ttbb)/\sigma(ttjj) \sim 11\%$
- JES  $\sim 9\%$
- JER, b-tagging  $\sim 6\%$

## OBS vs post-fit yields (BDT)



(similar plot for cut-based)



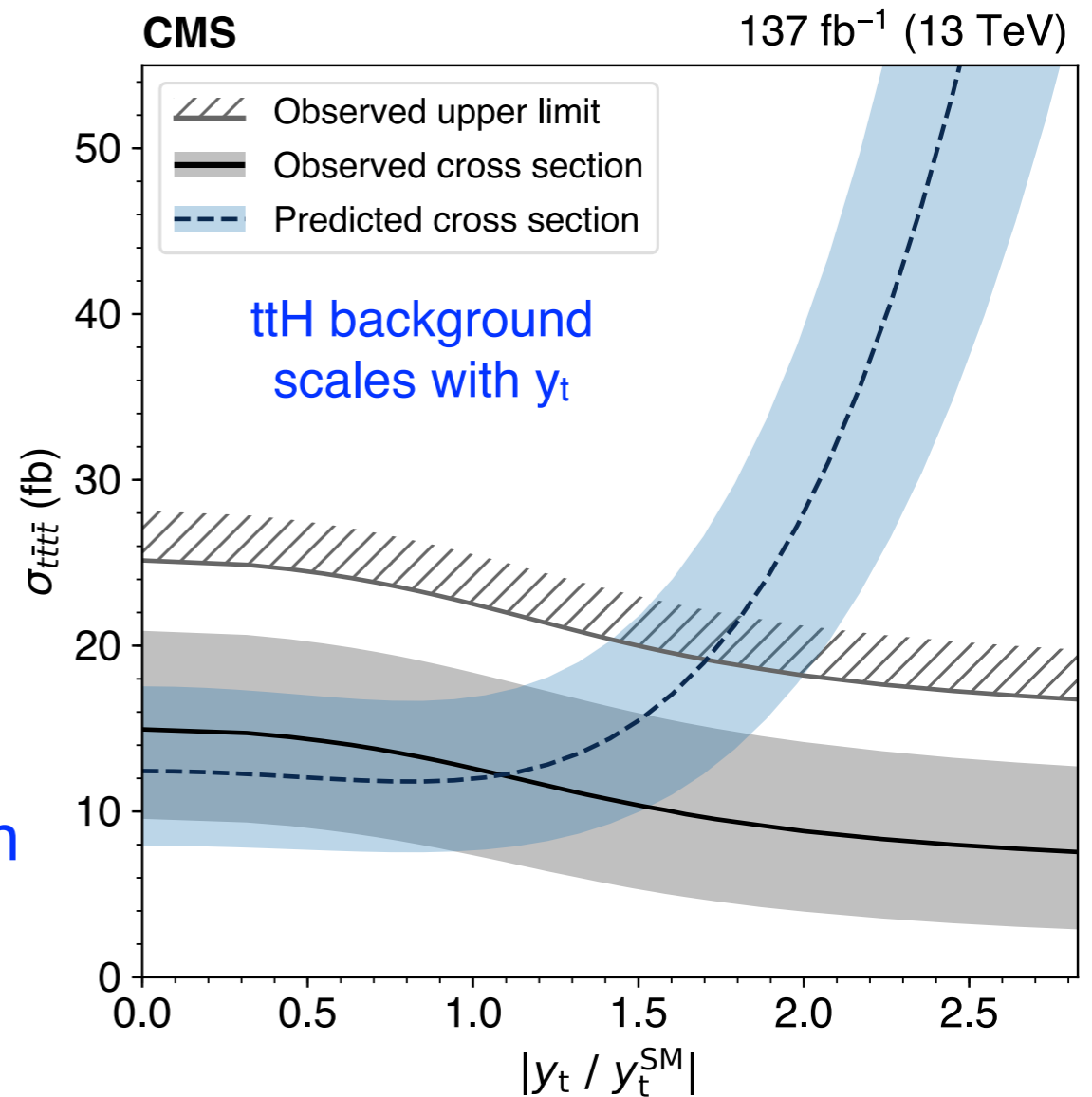
# tttt (same-sign dilepton and trilepton) /3

## Cross section:

- $\sigma(\text{tttt})_{\text{cut}} = 9.4 +6.2 -5.6 \text{ fb}$   
(1.7/2.5 $\sigma$  OBS/EXP significance w.r.t. bckgd-only)
- $\sigma(\text{tttt})_{\text{BDT}} = 12.6 +5.8 -5.2 \text{ fb}$   
(2.6/2.7 $\sigma$  OBS/EXP significance w.r.t. bckgd-only)
- 95% CL OBS upper limits = 20.0 / 22.5 fb
- 95% CL EXP upper limits = 9.4 / 8.5 fb
- vs  $\sigma(\text{tttt})_{\text{theory}} = 12.0 +2.2 -2.5 \text{ fb}$

The dependence of the (LO) SM cross section on the Yukawa coupling  $y_t$  is used to constrain  $y_t$

$$|y_t/y_t^{\text{SM}}| < 1.7 \text{ at } 95\% \text{ CL}$$

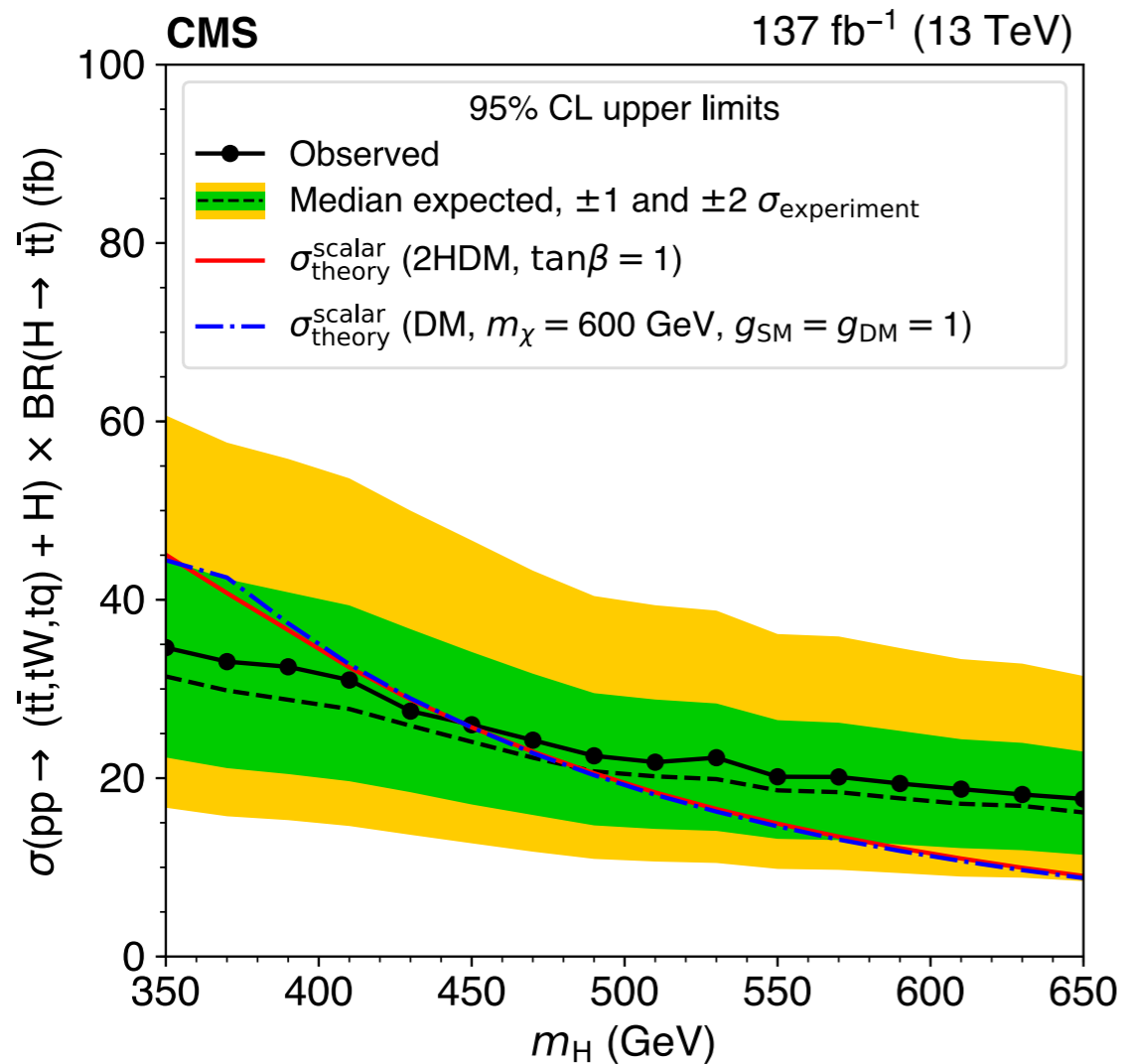




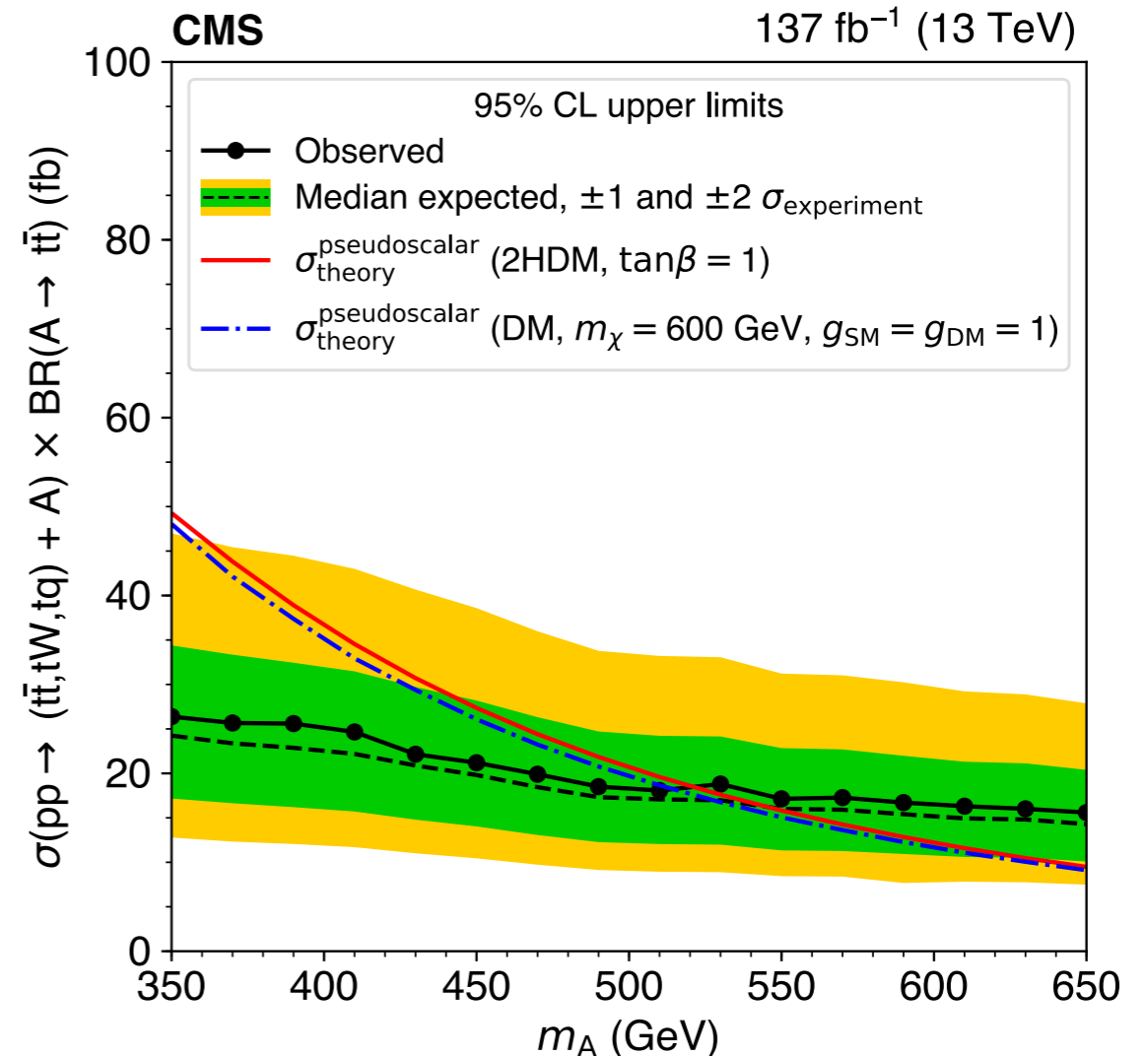
# tttt (same-sign dilepton and trilepton) /4

Constraints on new particles with  $m > 2m_t$

e.g. H and A from 2HDM, considering SM tttt as bckgd



$M_H > 470$  GeV  
at 95% CL



$M_A > 550$  GeV  
at 95% CL

# Conclusions

## CMS measures the production of $t\bar{t}$ + generic/HF jets:

- $t\bar{t}j$  cross section as expected,  $t\bar{t}b\bar{b}$  cross section  $\approx 30\%$  higher
- increased precision on  $t\bar{t}b\bar{b}$  production with latest measurement
- need to reduce stat. uncert. for differential measurements

Getting close to observe  $t\bar{t}t\bar{t}$  production ( $2.6\sigma$  OBS significance)

**CMS**  
*Preliminary*

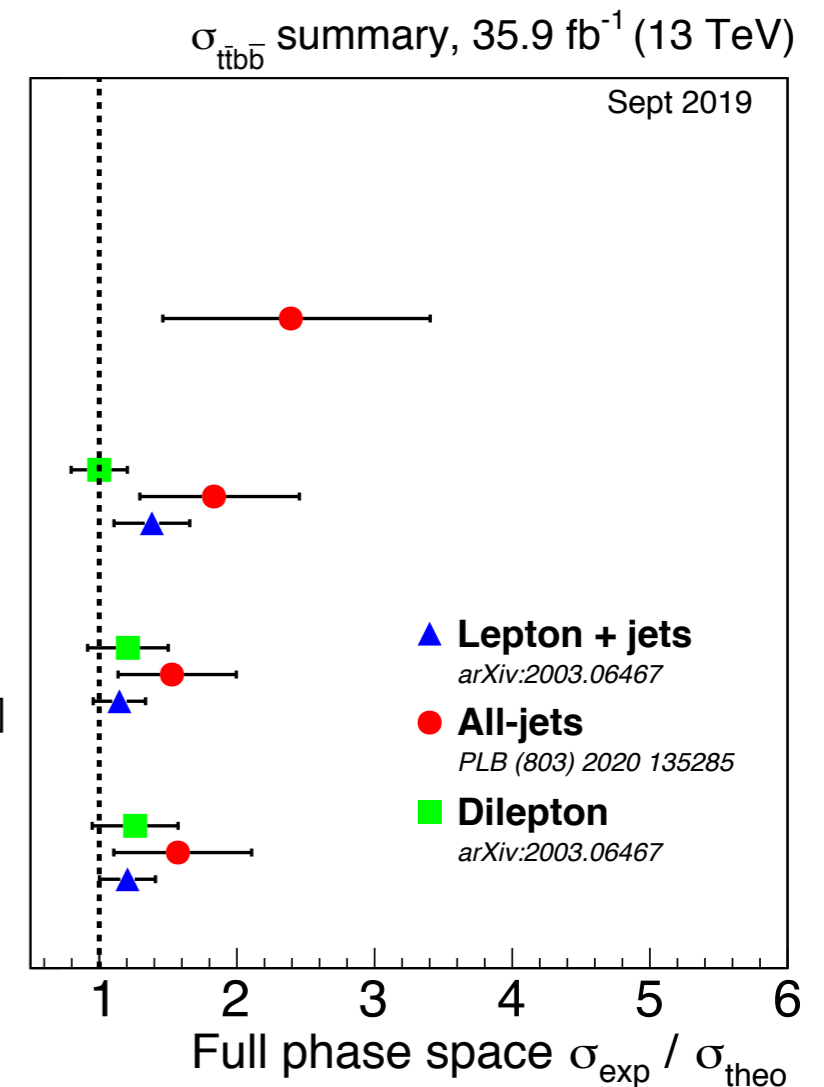
Reference for  $\sigma_{\text{theo}}$

MG5\_aMC@NLO +  
PYTHIA8 4FS

POWHEG +  
HERWIG++

MG5\_aMC@NLO +  
PYTHIA8 5FS [FxFx]

POWHEG +  
PYTHIA8



# Backup



# ttjj + ttbb (lepton + dilepton) /4

## systematics

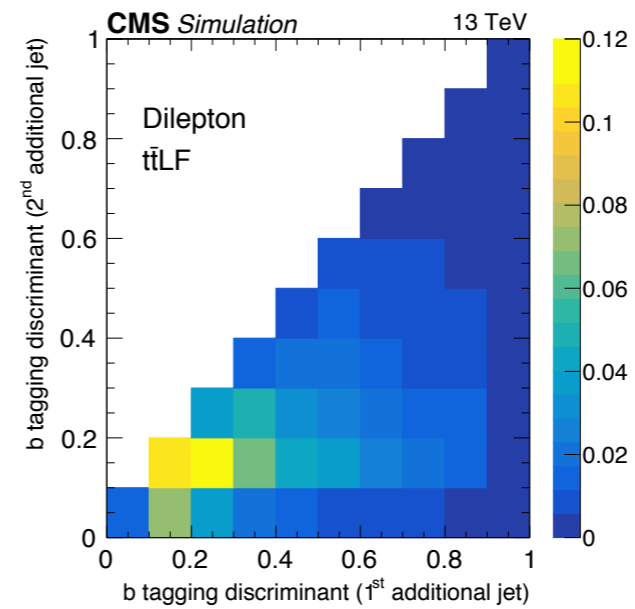
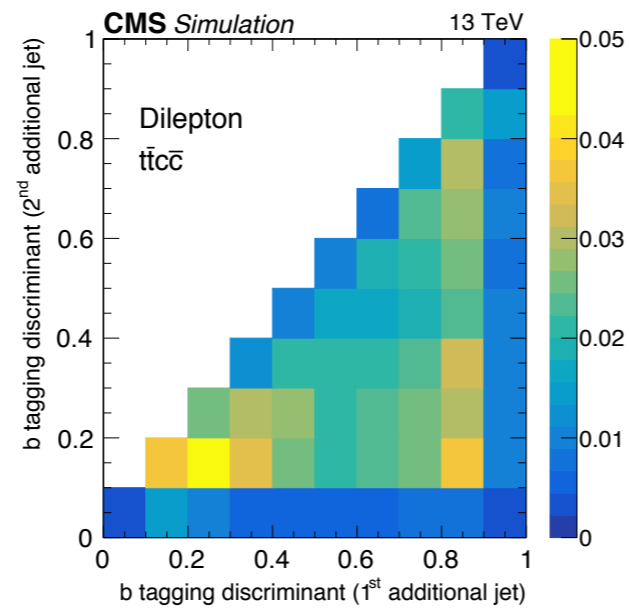
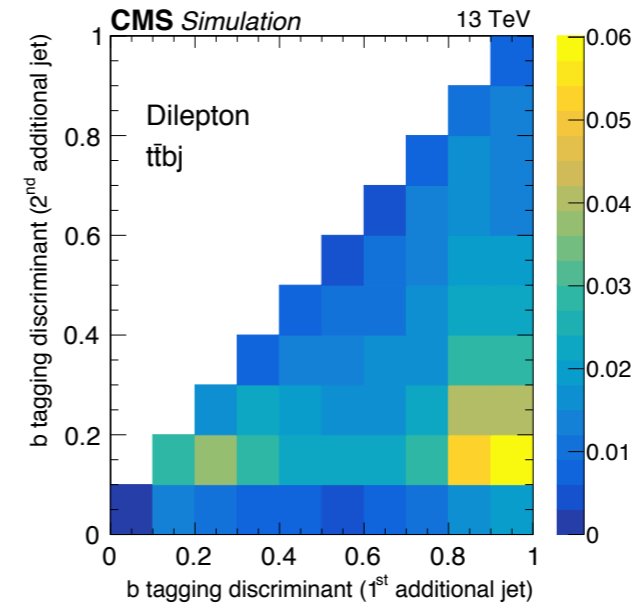
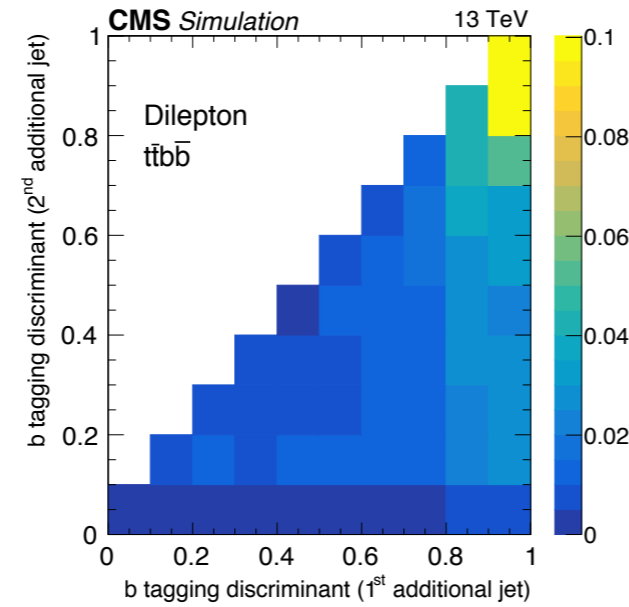
Source	$R_{\text{tt}\bar{b}\bar{b}/\text{ttjj}}^{\text{VPS}} [\%]$		$\sigma_{\text{ttjj}}^{\text{VPS}} [\%]$	
	Dilepton	Lepton+jets	Dilepton	Lepton+jets
Lepton uncertainties				
Trigger	<0.1	0.2	1.0	0.5
Lepton identification	0.6	0.2	1.1	1.3
Lepton energy scale	—	<0.1	—	0.1
Jet uncertainties				
Jet energy resolution (JER)	0.4	0.3	0.3	0.7
Jet energy scale (JES)	1.5	1.2	2.9	3.6
b tagging uncertainties				
c-flavor btag (lin.)	2.2	2.0	1.0	0.3
c-flavor btag (quad.)	0.7	1.2	0.3	0.2
Heavy-flavor btag	4.0	0.1	0.5	0.9
Heavy-flavor btag (lin.)	0.9	0.4	1.5	0.5
Heavy-flavor btag (quad.)	2.0	0.3	1.5	0.8
Light-flavor btag	4.9	0.9	5.5	4.9
Light-flavor btag (lin.)	0.1	0.2	0.3	1.1
Light-flavor btag (quad.)	0.7	0.7	0.1	1.4
Theoretical uncertainties				
Initial-state radiation (ISR)	1.0	2.2	2.5	1.2
Final-state radiation (FSR)	0.8	0.7	2.5	5.9
ME-PS matching	0.5	<0.1	1.8	1.9
Underlying event tune (UE)	1.5	1.5	0.4	1.4
$\mu_F/\mu_R$ scales (ME)	0.1	0.4	0.1	1.4
top- $p_T$	0.2	0.4	1.6	0.3
Ratio $R_{\text{ttbj}/\text{tt}\bar{b}\bar{b}}^{\text{MC}}$	1.4	0.2	1.3	0.7
Other uncertainties				
Pileup	0.7	0.2	1.3	0.1
Backgrounds	0.3	2.0	0.7	1.2
Simulated sample size	1.5	2.8	0.1	2.2
Luminosity	0.2	0.5	2.6	3.1
Total	8.0	5.5	8.8	10.0

## measurements

	$R_{\text{tt}\bar{b}\bar{b}/\text{ttjj}}$	$\sigma_{\text{ttjj}} [\text{pb}]$	$\sigma_{\text{tt}\bar{b}\bar{b}} [\text{pb}]$
Dilepton channel (VPS)			
POWHEG + PYTHIA8	$0.013 \pm 0.002$	$2.41 \pm 0.21$	$0.032 \pm 0.004$
Measurement	$0.017 \pm 0.001 \pm 0.001$	$2.36 \pm 0.02 \pm 0.20$	$0.040 \pm 0.002 \pm 0.005$
Dilepton channel (FPS)			
POWHEG + PYTHIA8	$0.014 \pm 0.003$	$163 \pm 21$	$2.3 \pm 0.4$
MG_aMC@NLO + PYTHIA8 5FS [FxFx]	$0.015 \pm 0.003$	$159 \pm 25$	$2.4 \pm 0.4$
POWHEG + HERWIG++	$0.011 \pm 0.002$	$170 \pm 25$	$1.9 \pm 0.3$
Measurement	$0.018 \pm 0.001 \pm 0.002$	$159 \pm 1 \pm 15$	$2.9 \pm 0.1 \pm 0.5$
Lepton+jets channel (VPS)			
POWHEG + PYTHIA8	$0.017 \pm 0.002$	$30.5 \pm 3.0$	$0.52 \pm 0.06$
Measurement	$0.020 \pm 0.001 \pm 0.001$	$31.0 \pm 0.2 \pm 2.9$	$0.62 \pm 0.03 \pm 0.07$
Lepton+jets channel (FPS)			
POWHEG + PYTHIA8	$0.013 \pm 0.002$	$290 \pm 29$	$3.9 \pm 0.4$
MG_aMC@NLO + PYTHIA8 5FS [FxFx]	$0.014 \pm 0.003$	$280 \pm 40$	$4.1 \pm 0.4$
POWHEG + HERWIG++	$0.011 \pm 0.002$	$321 \pm 36$	$3.4 \pm 0.5$
Measurement	$0.016 \pm 0.001 \pm 0.001$	$292 \pm 1 \pm 29$	$4.7 \pm 0.2 \pm 0.6$

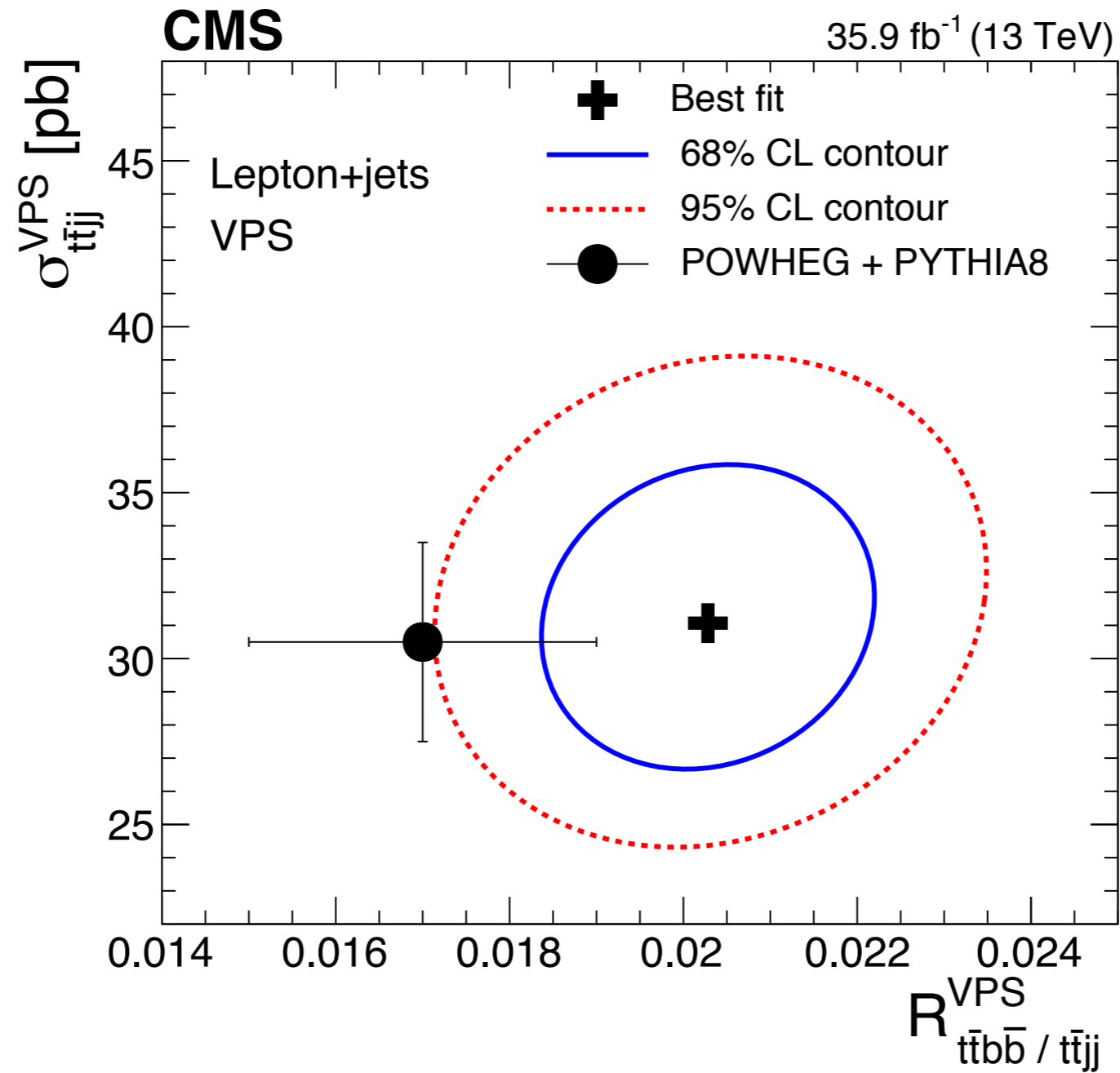


# $ttjj + ttbb$ (lepton + dilepton) /5



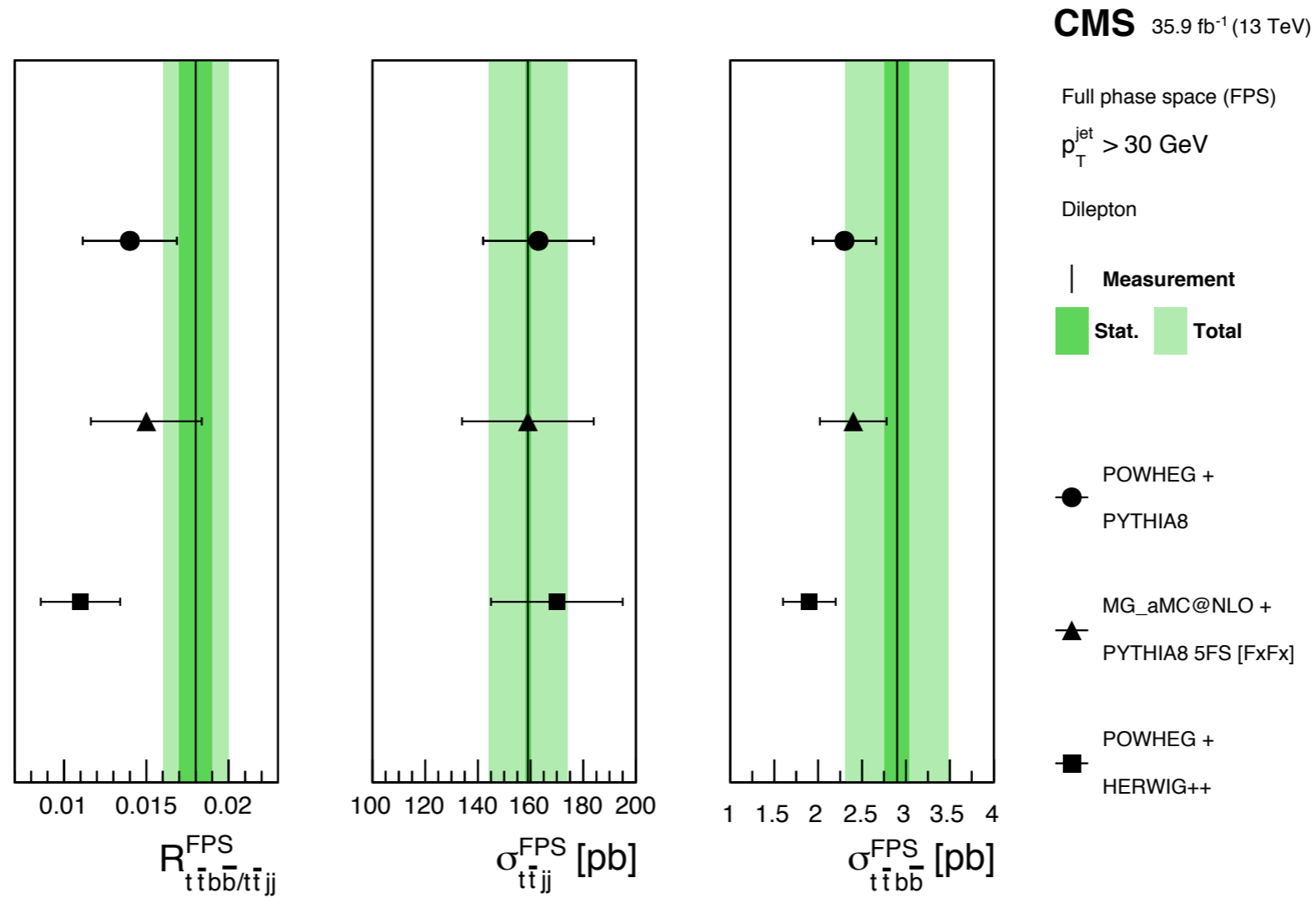


# $ttjj + ttbb$ (lepton + dilepton) /6





# ttjj + ttbb (lepton + dilepton) /7





# ttbb (all-had) /4

## systematics

Source	FPS PI (%)	FPS PB (%)
Simulated sample size	+15 -11	+15 -11
Quark-gluon likelihood	+13 -8	+13 -8
b tagging of b quark	$\pm 10$	$\pm 10$
JES and JER	+5.1 -5.2	+5.0 -5.4
Integrated luminosity	+2.8 -2.2	+2.4 -2.2
Trigger efficiency	+2.6 -2.1	+2.5 -2.2
Pileup	+2.3 -2.0	+2.2 -1.9
$\mu_R$ and $\mu_F$ scales	+13 -9	+13 -9
Parton shower scale	+11 -8	+11 -8
UE tune	+9.0 -5.3	+9.0 -5.2
Colour reconnection	$\pm 7.2$	$\pm 7.1$
Shower matching ( $h_{\text{damp}}$ )	+4.3 -2.8	+3.8 -2.7
$t\bar{t}c\bar{c}$ normalization	+3.2 -4.4	+2.9 -4.5
Modelling of $p_T$ of top quark	$\pm 2.5$	$\pm 2.4$
PDFs	+2.2 -2.0	+2.2 -2.0
Total	+28 -23	+28 -23



# tttt (single lepton + opposite-sign dilepton) /4

## systematics

Systematic uncertainty	Normalization	Shape
Integrated luminosity	X	
Pileup modeling	X	X
Lepton reconstruction and identification	X	
Jet energy corrections	X	X
b tagging	X	X
Ren. and fact. scales	X	X
PS scales	X	
ME-PS matching	X	
UE	X	
Jet multiplicity correction	X	
Parton distribution functions	X	X
Top quark $p_T$ reweighting		X
Heavy-flavor reweighting	X	X
Rare process	X	



# tttt (same-sign dilepton and trilepton) /5

## systematics

Source	Uncertainty (%)	Impact on $\sigma(\bar{t}t\bar{t})$ (%)
Integrated luminosity	2.3–2.5	2
Pileup	0–5	1
Trigger efficiency	2–7	2
Lepton selection	2–10	2
Jet energy scale	1–15	9
Jet energy resolution	1–10	6
b tagging	1–15	6
Size of simulated sample	1–25	<1
Scale and PDF variations †	10–15	2
ISR/FSR (signal) †	5–15	2
tH (normalization) †	25	5
Rare, X $\gamma$ , tV (norm.) †	11–20	<1
tZ, tW (norm.) †	40	3–4
Charge misidentification †	20	<1
Nonprompt leptons †	30–60	3
$N_{\text{jets}}^{\text{ISR/FSR}}$	1–30	2
$\sigma(\bar{t}t\bar{b}\bar{b})/\sigma(\bar{t}t\bar{j}j)$ †	35	11

## yields

	SM background	t $\bar{t}\bar{t}$	Total	Observed
CRZ	102 ± 12	1.11 ± 0.43	103 ± 12	104
SR1	3.95 ± 0.96	< 0.01	3.96 ± 0.96	4
SR2	14.2 ± 1.8	0.01 ± 0.01	14.2 ± 1.8	19
SR3	25.5 ± 3.5	0.04 ± 0.03	25.6 ± 3.5	19
SR4	34.0 ± 4.0	0.08 ± 0.05	34.0 ± 4.0	33
SR5	36.7 ± 4.0	0.15 ± 0.07	36.8 ± 4.0	36
SR6	39.8 ± 4.2	0.23 ± 0.12	40.0 ± 4.2	44
SR7	40.3 ± 3.7	0.31 ± 0.16	40.6 ± 3.8	41
SR8	47.3 ± 4.3	0.72 ± 0.28	48.0 ± 4.3	46
SR9	58.5 ± 5.2	1.18 ± 0.46	59.7 ± 5.2	48
SR10	52.1 ± 4.3	1.91 ± 0.74	54.1 ± 4.2	61
SR11	43.0 ± 3.5	3.0 ± 1.2	46.0 ± 3.5	62
SR12	32.1 ± 3.0	3.7 ± 1.4	35.8 ± 2.9	40
SR13	16.7 ± 1.6	4.3 ± 1.6	21.0 ± 2.0	15
SR14	10.1 ± 1.2	4.2 ± 1.6	14.3 ± 1.8	16
SR15	5.03 ± 0.77	4.1 ± 1.5	9.1 ± 1.6	4
SR16	2.49 ± 0.61	3.4 ± 1.3	5.9 ± 1.3	7
SR17	0.57 ± 0.36	1.08 ± 0.42	1.65 ± 0.50	3

BDT analysis



# $t\bar{t}t\bar{t}$ (same-sign dilepton and trilepton) /6

