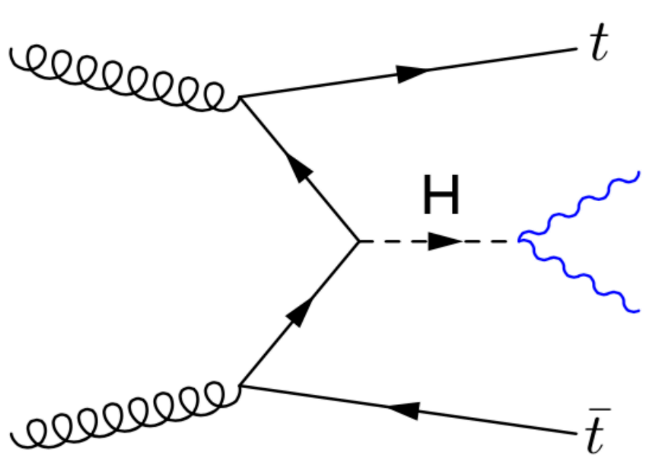
What is wrong with multileptons + b-jets?

Elizaveta Shabalina
University of Göttingen

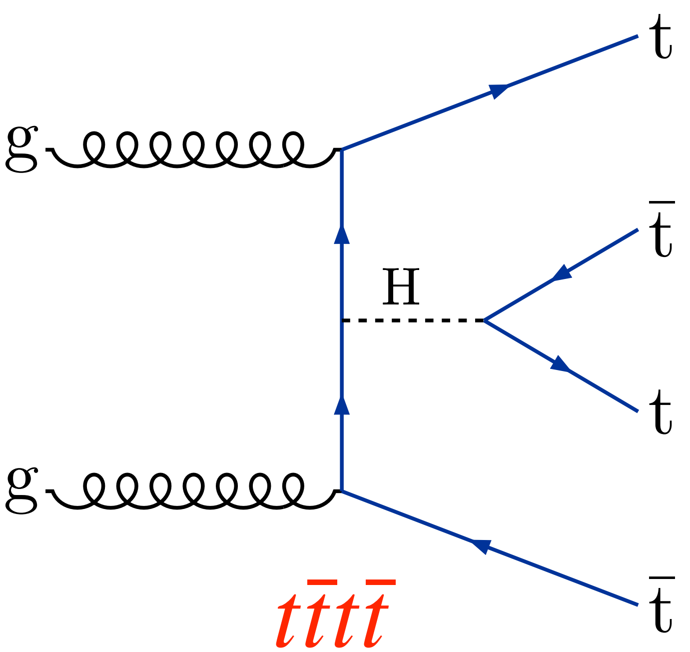
Didar Dobur
University of Ghent

Introduction

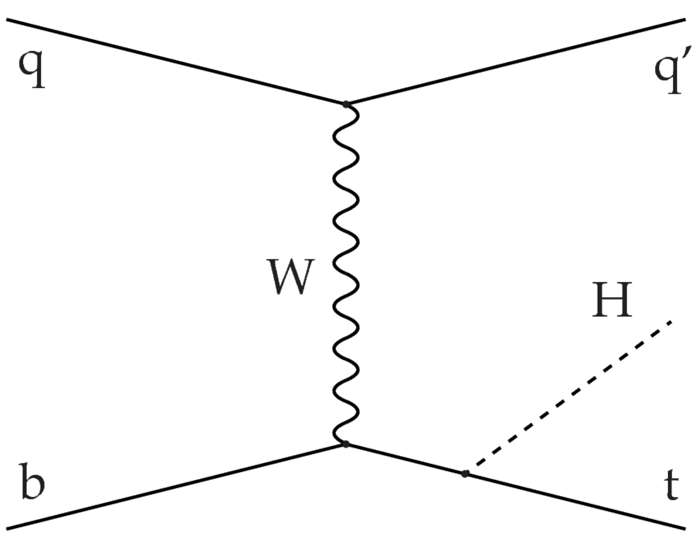
- Multilepton (e,μ) final states with b-jets:
  - 2-leptons with the same electric charge (2LSS)
  - 3-leptons



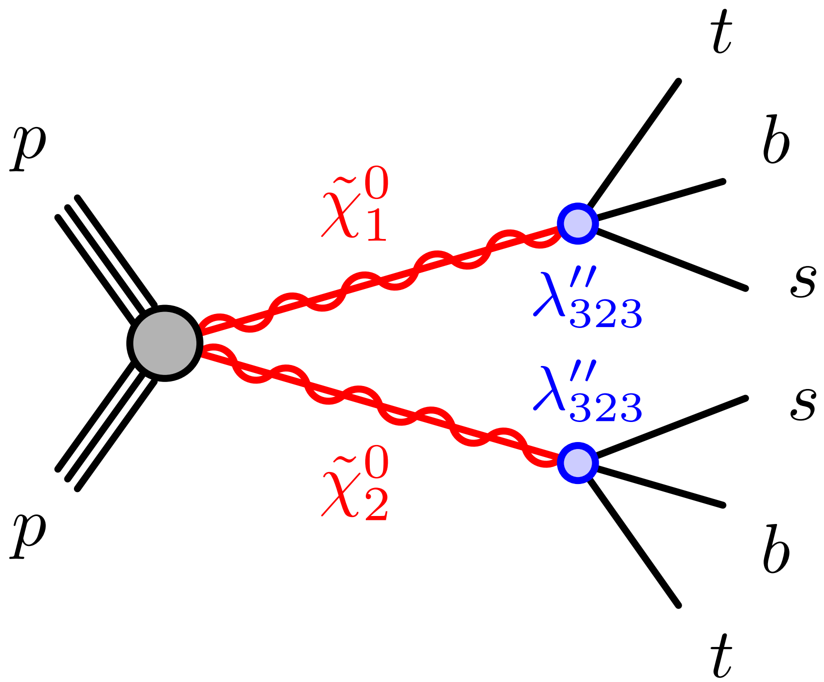
$t\bar{t}H$



$t\bar{t}t\bar{t}$

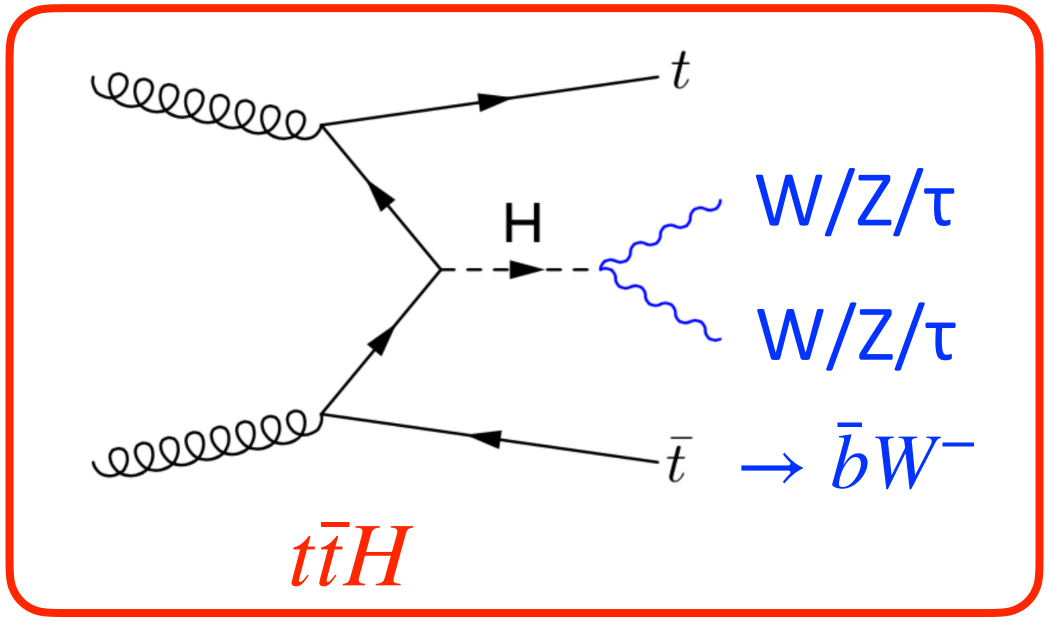


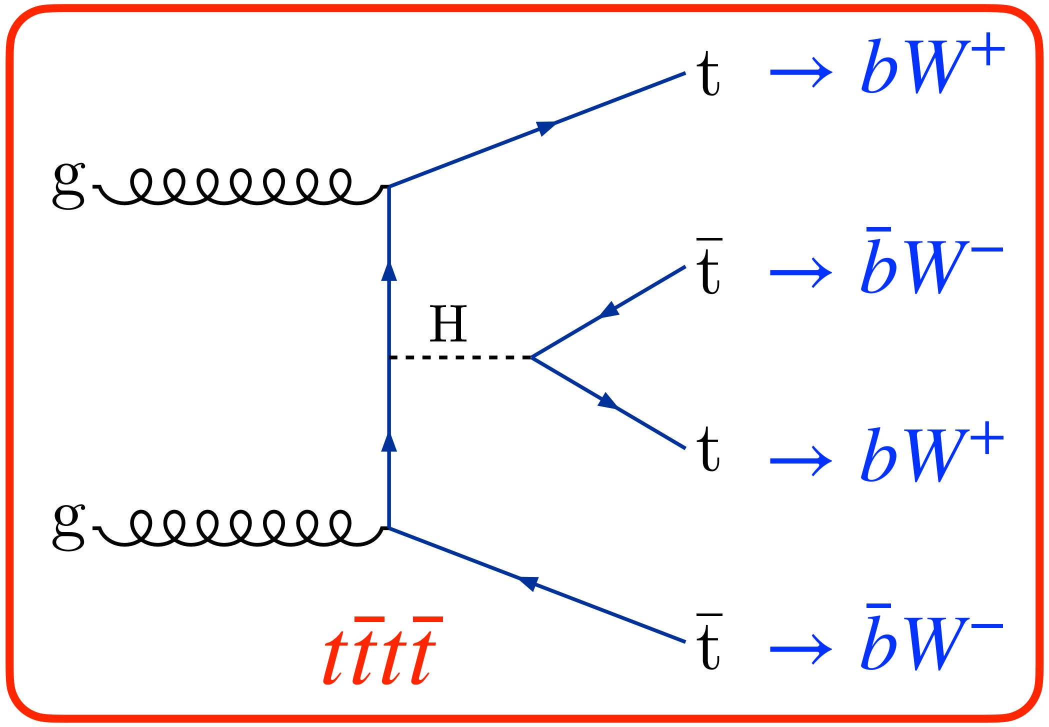
tHq

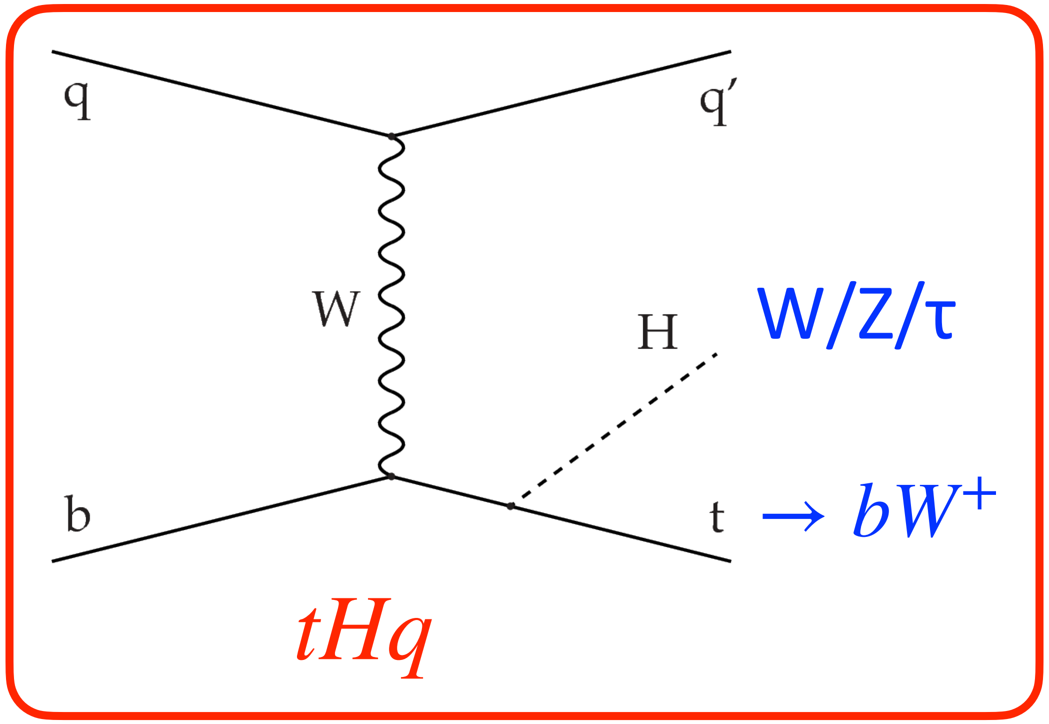


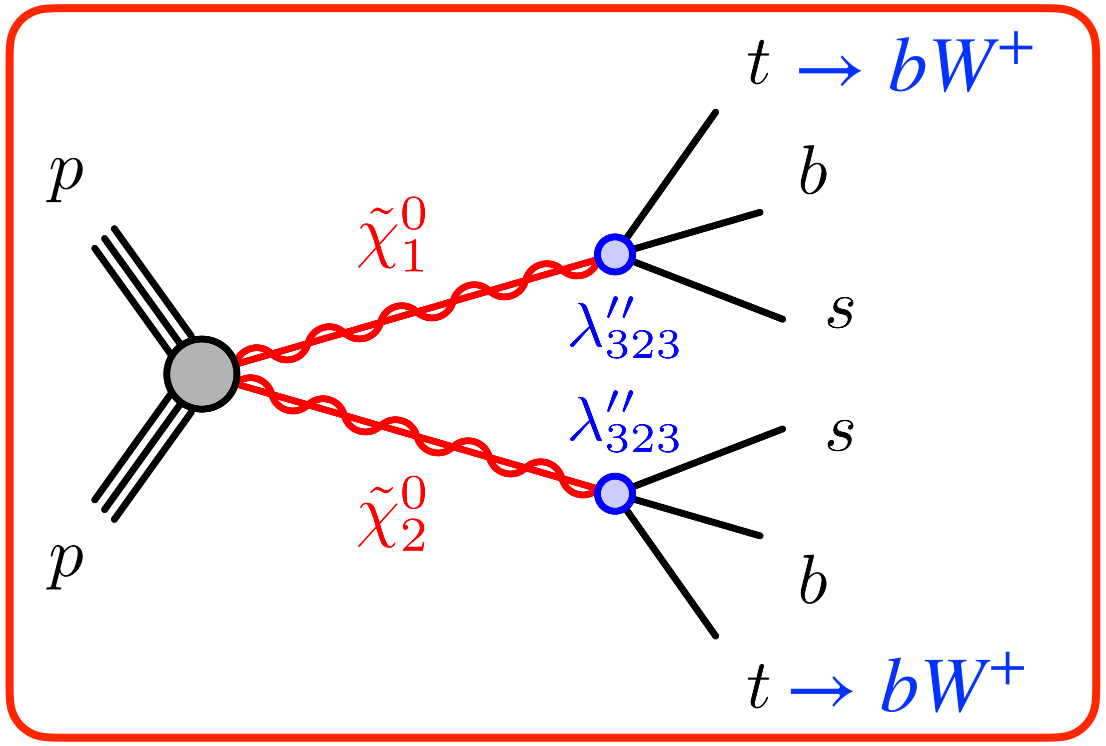
Introduction

- Multilepton (e,μ) final states with b-jets:
  - 2-leptons with the same electric charge (2LSS)
  - 3-leptons







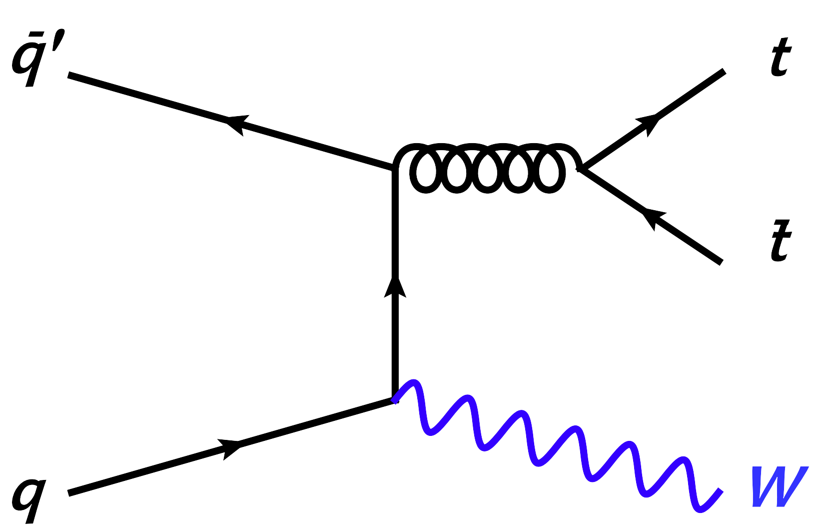


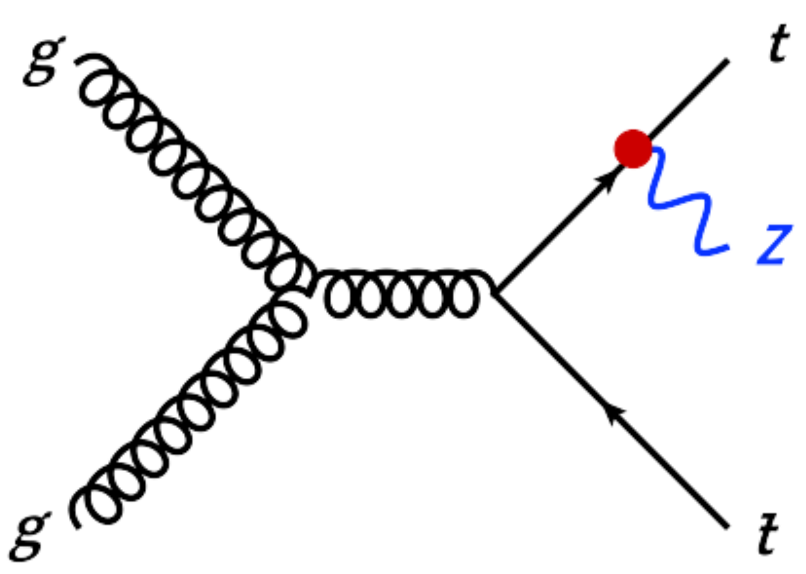
Common features

- Multilepton selection
  - Advantage: suppresses SM background
  - Disadvantage: low branching fractions of W,Z

- 2LSS
  - ttW is the main background
  - ttZ, dibosons
  - charge mis-identification (electron)
  - fake leptons

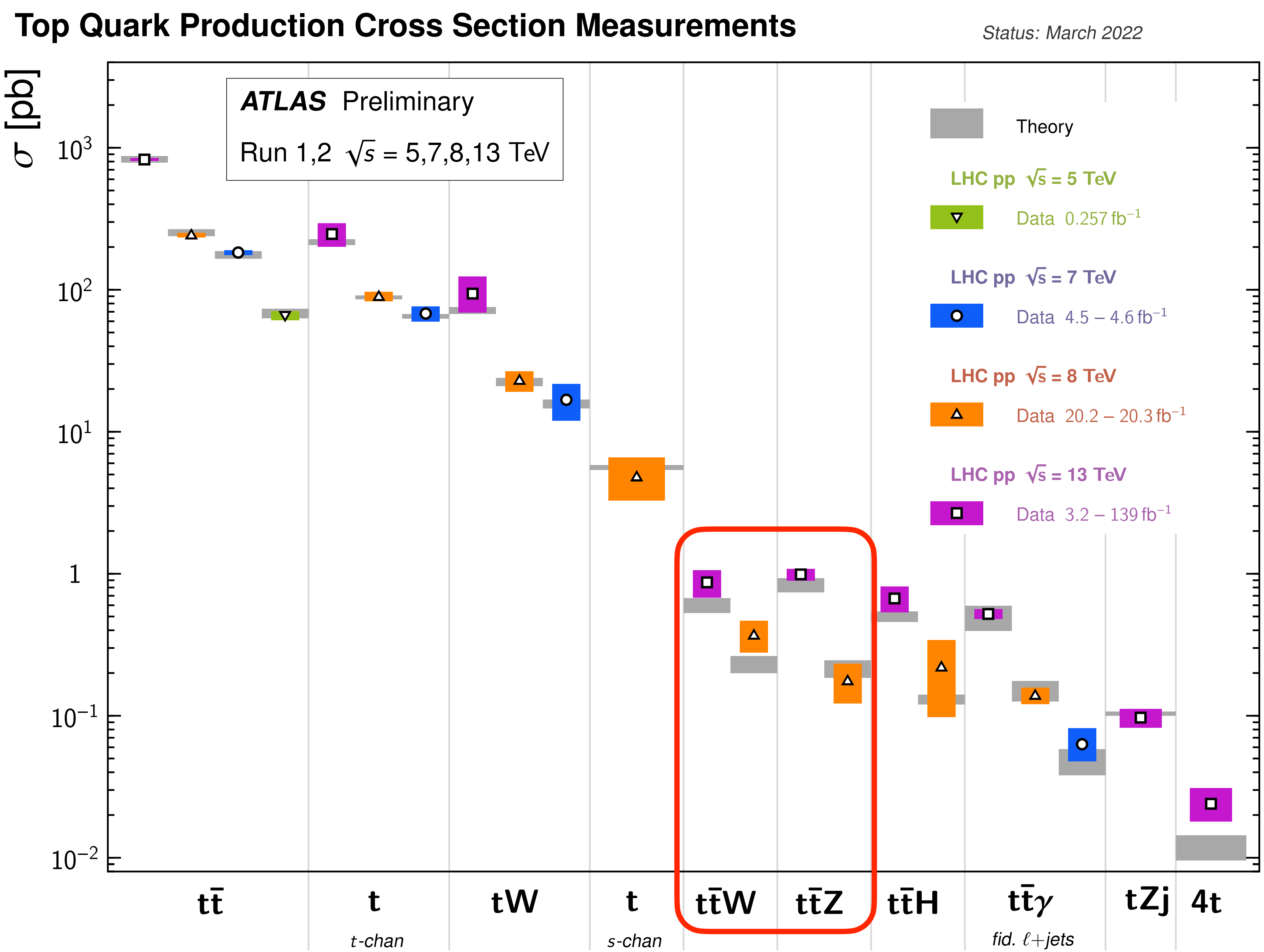
- 3L
  - lower background than in 2LSS
  - ttW, ttZ, dibosons
  - fake leptons





GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

tt+X status



- $t\bar{t}Z$ and $t\bar{t}W$ are quite rare processes by themselves
- Measurements with full run 2 data set started to appear last year

ttZ measurement

ttZ measurement

- Selection
  - 3 or 4 leptons
  - Z candidate
  - CMS: $N_{jets} \geq 2$, $N_b \geq 0$
  - ATLAS: $N_{jets} \geq 3$ with varied requirements on tightness and number of b-jets for inclusive and differential measurements
- Backgrounds
  - Diboson+b-jets
  - non-prompt leptons
- Diboson model validated in CRs included in the fit

Main focus of ATLAS are the differential distributions

ttZ signal

- nominal: MG5_aMC@NLO + Pythia
- alternative:
  - MG5_aMC@NLO + Herwig
  - Sherpa NLO inclusive
  - Sherpa multilevel with 1 parton
  - Theory fixed order calculation at NLO, NLO+NLL, nNLO (JHEP 08 (2019) 039)

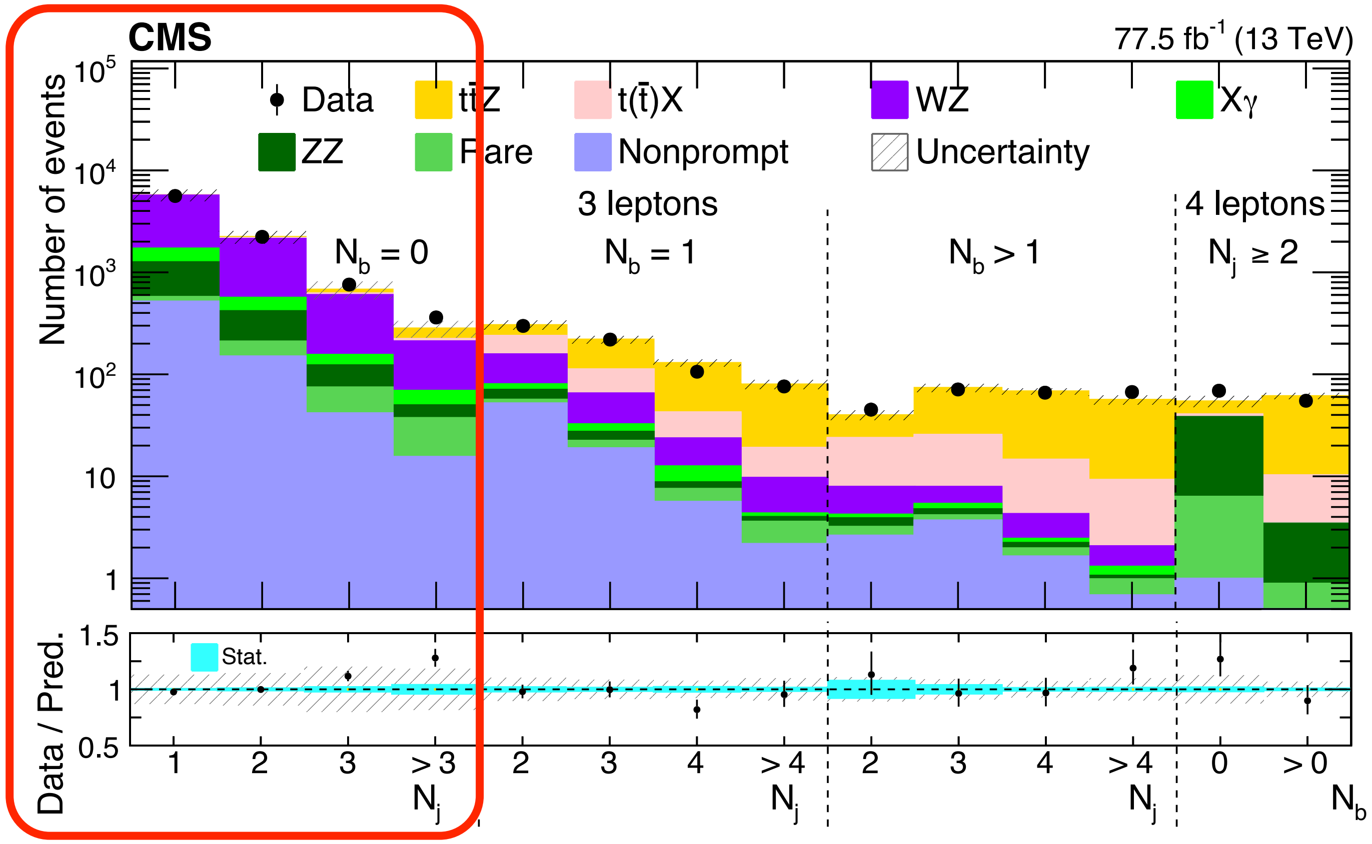
$$\sigma = 0.88^{+0.09}_{-010}\text{pb}$$

Extrapolated from YR4 to include off-shell effects

ttZ inclusive

- Simultaneous profile likelihood fit in multiple signal/control regions

JHEP 03 (2020) 056



$$\mu = 1.13 \pm 0.06(\text{stat}) \pm 0.07(\text{syst})\ \text{pb}$$

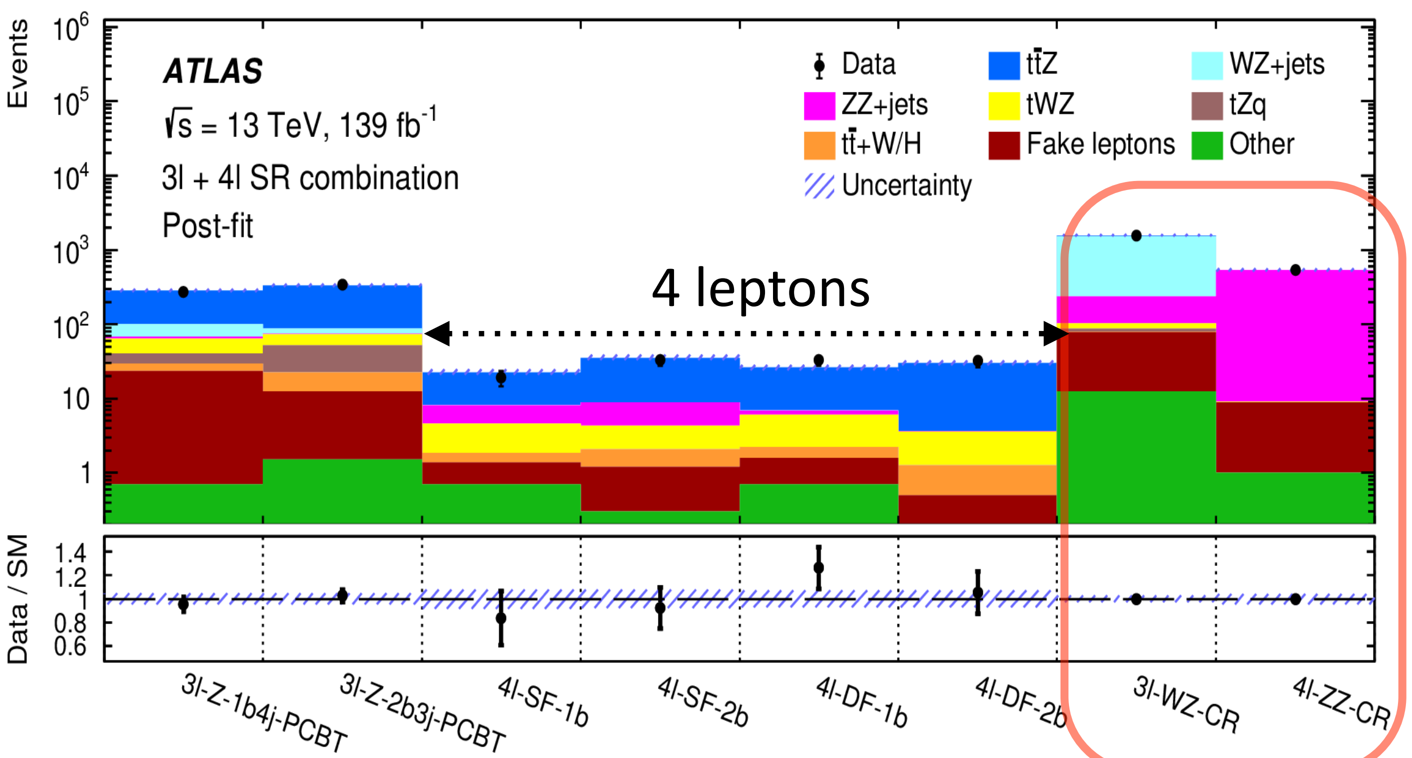
8% precision

$$\sigma = 0.95 \pm 0.05(\text{stat}) \pm 0.06(\text{syst})\ \text{pb}$$

Main systematic uncertainties

- lepton identification (4%)
- WZ (3%) and t(t)X (3%)

Eur. Phys. J. C 81 (2021) 737



$$\mu = 1.19 \pm 0.06(\text{stat}) \pm 0.10(\text{syst})\ \text{pb}$$

10% precision

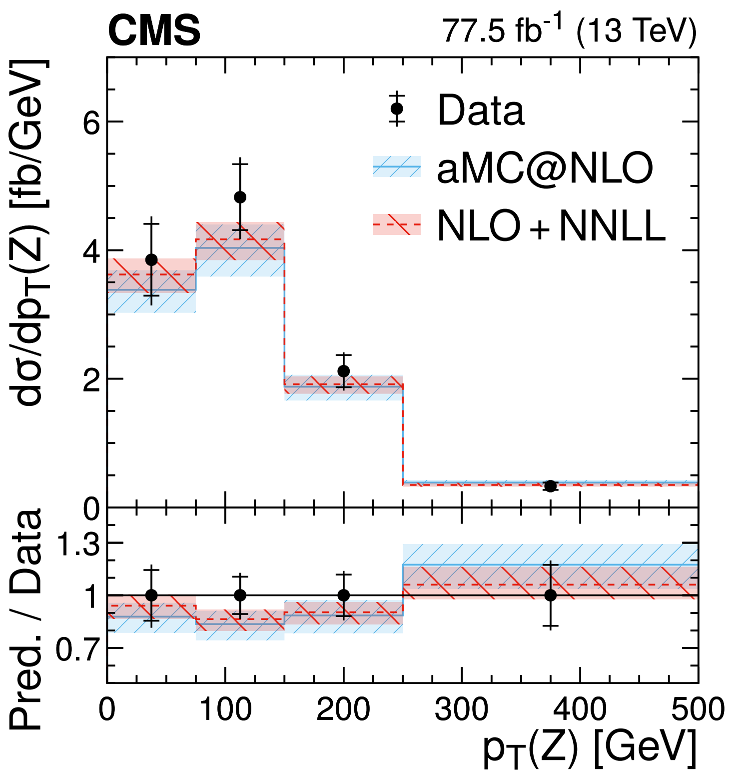
$$\sigma = 0.99 \pm 0.05(\text{stat}) \pm 0.08(\text{syst})\ \text{pb}$$

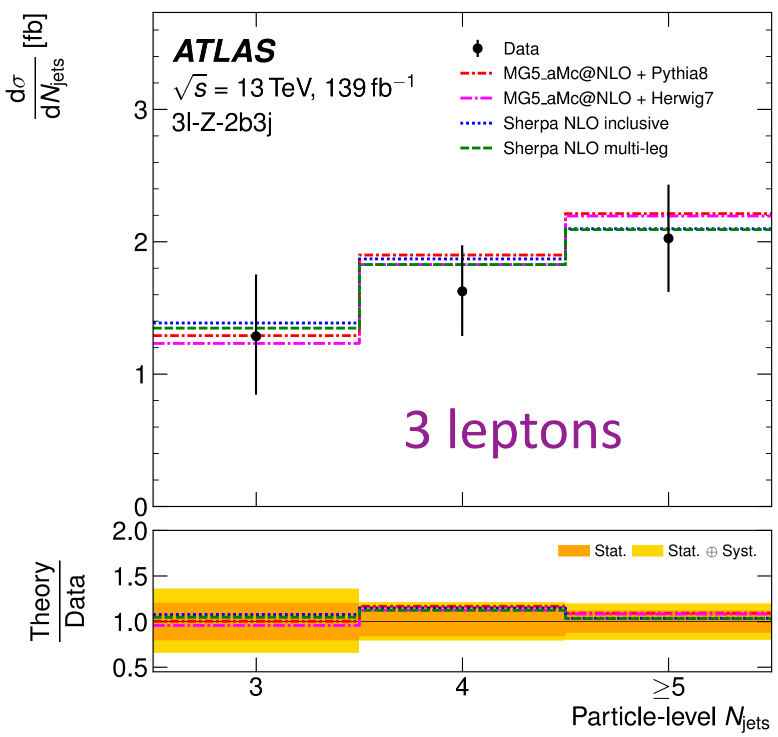
Main systematic uncertainties

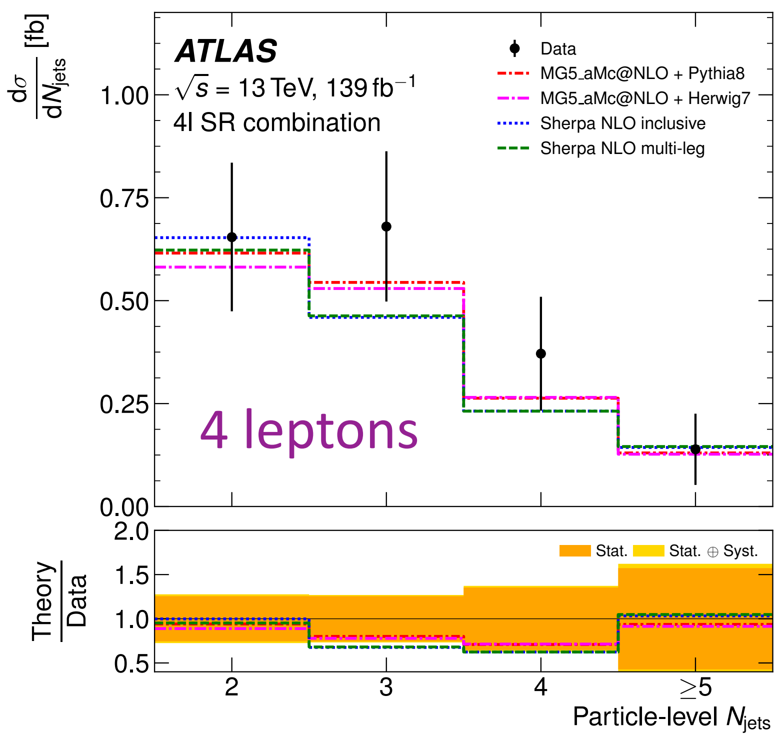
- ttZ parton shower model (3.1%)
- tWZ model, b-tagging (2.9% each)
- flavour tagging (2.8%)

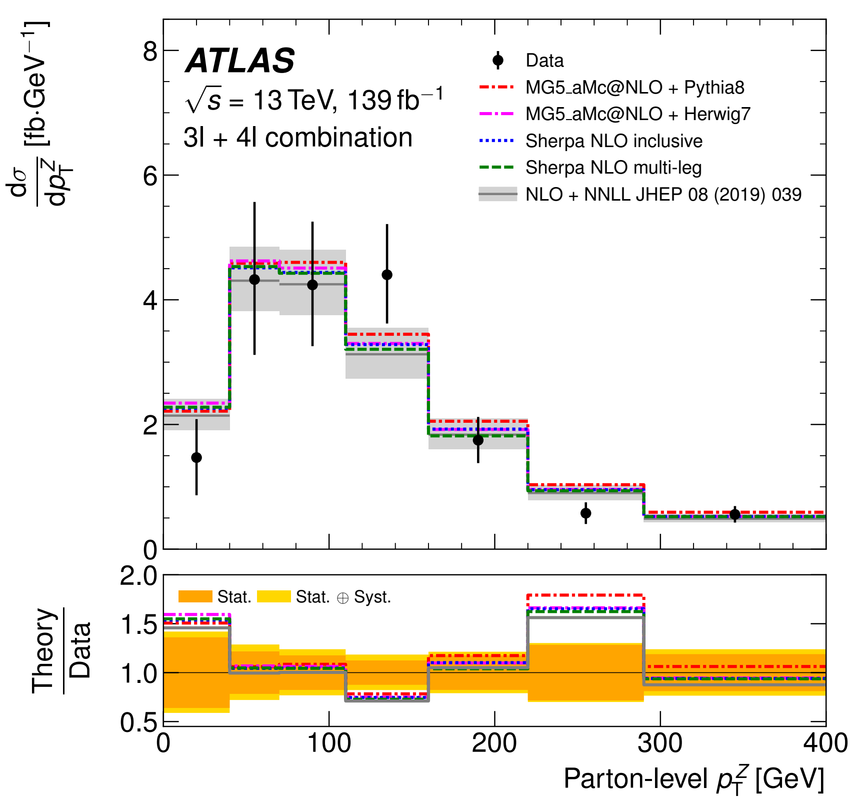
ttZ: going differential

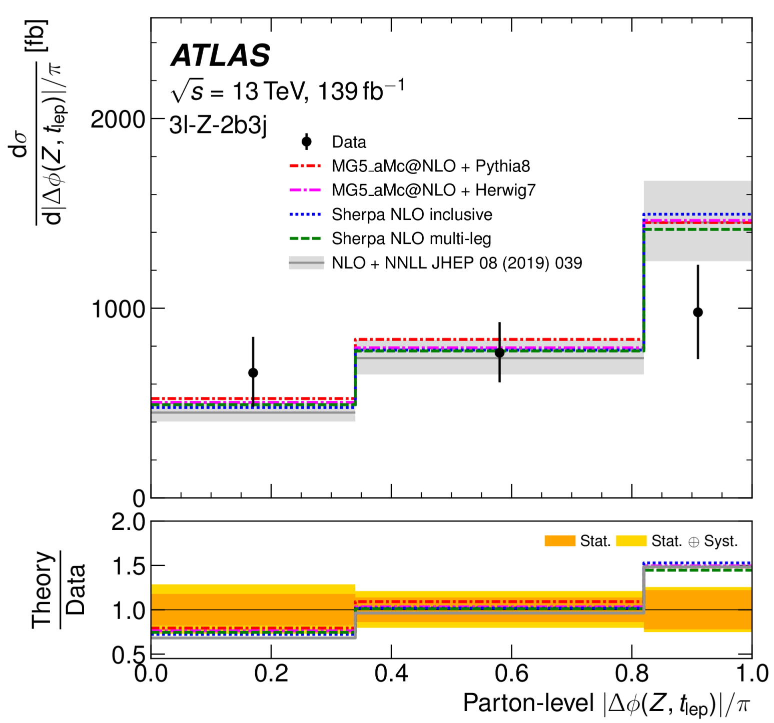
- Differential measurements to validate and improve ttZ MC models









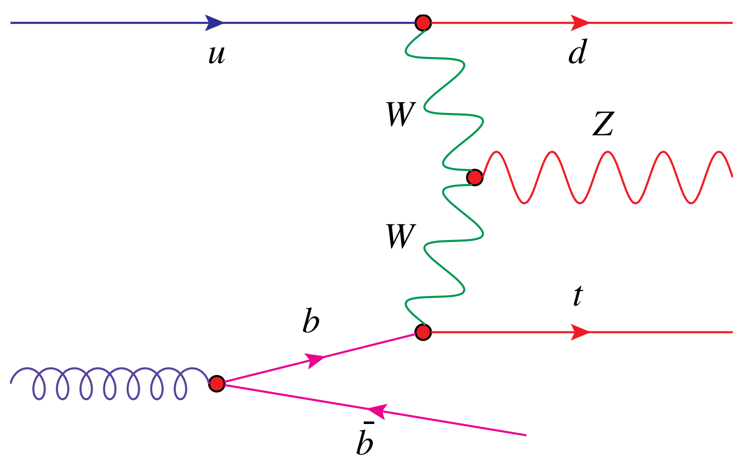


Good agreement with all tested MC models and theory calculation within uncertainties

Eur. Phys. J. C 81 (2021) 737

ttZ in tZq measurement

JHEP 02 (2022) 107



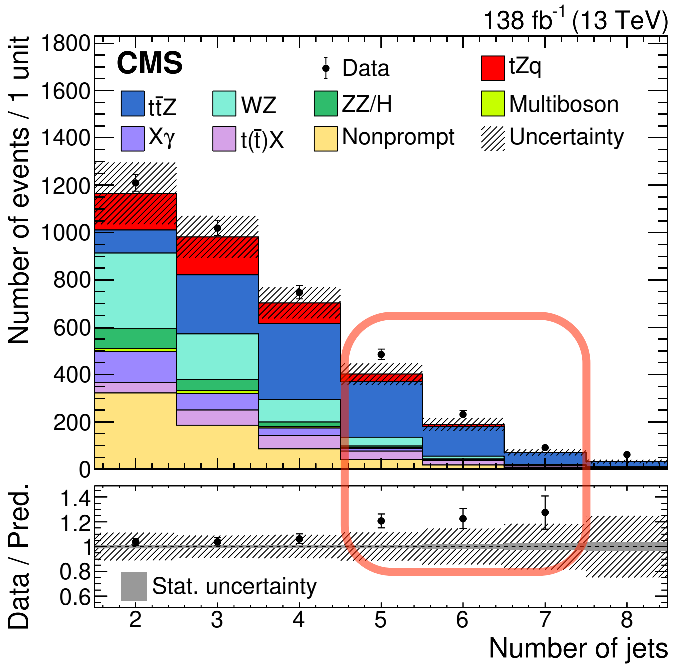
forward jet

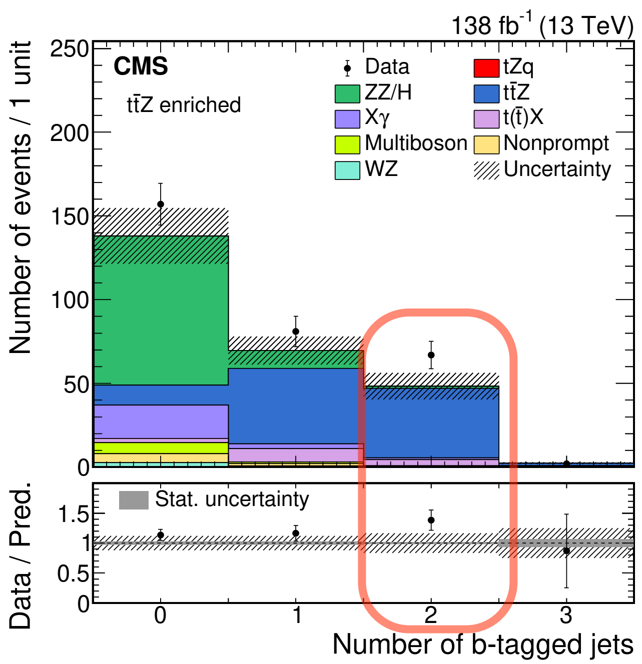
$Z(\ell\ell)$

$b\ell\nu$

- Selection
  - 3 leptons, Z candidate ≥2 jets, ≥1 b-jet
- 3 regions
  - 1 b-jet, 2-3 jets; ≥4 jets; ≥2 b-jets

- small deficit of ttZ compared to data in bins dominated by ttZ, especially at high jet multiplicity

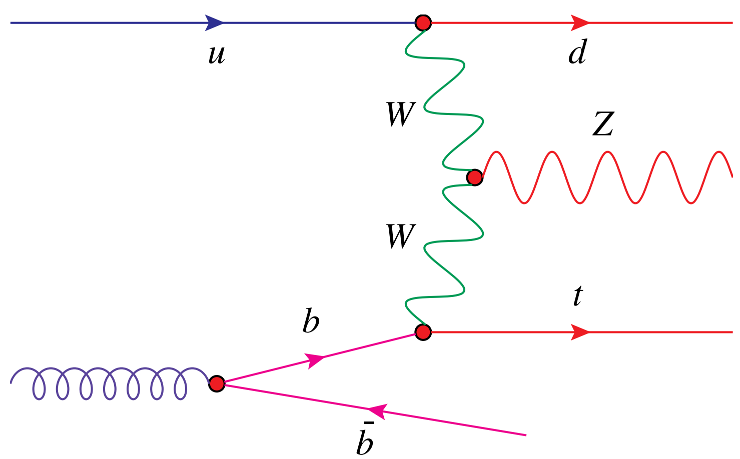




ttZ in tZq measurement

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

JHEP 02 (2022) 107

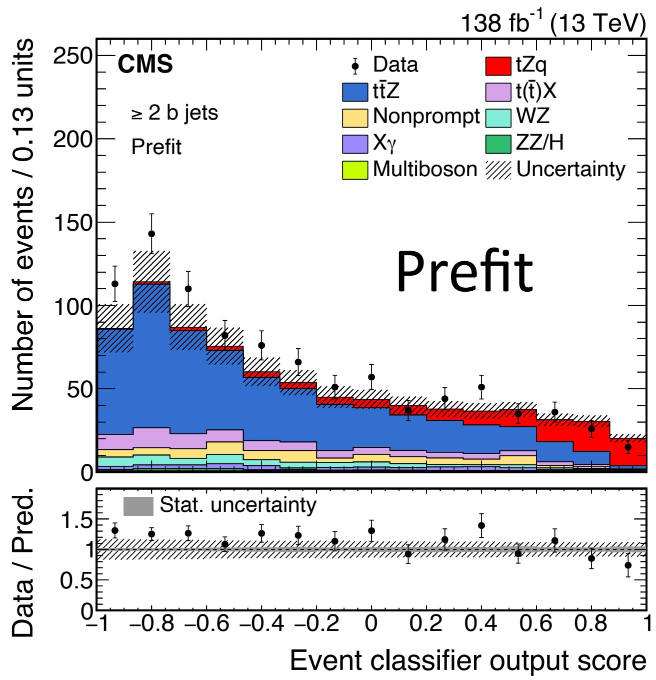


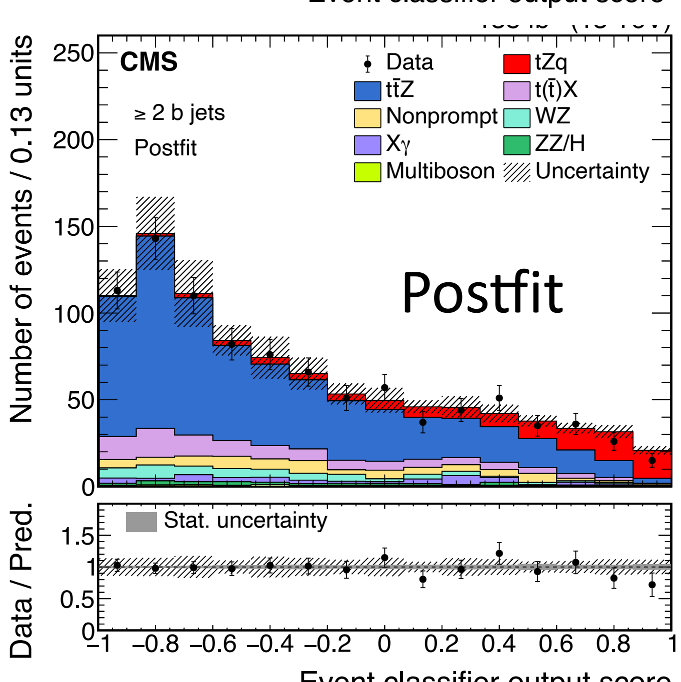
forward jet

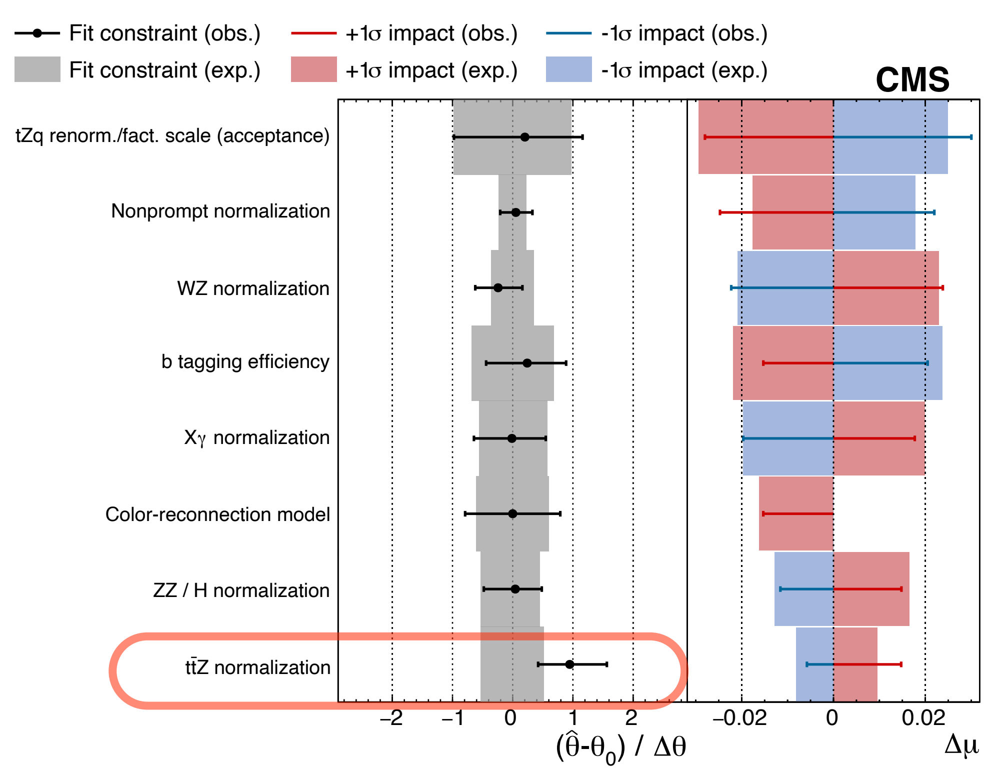
Z(ℓℓ)

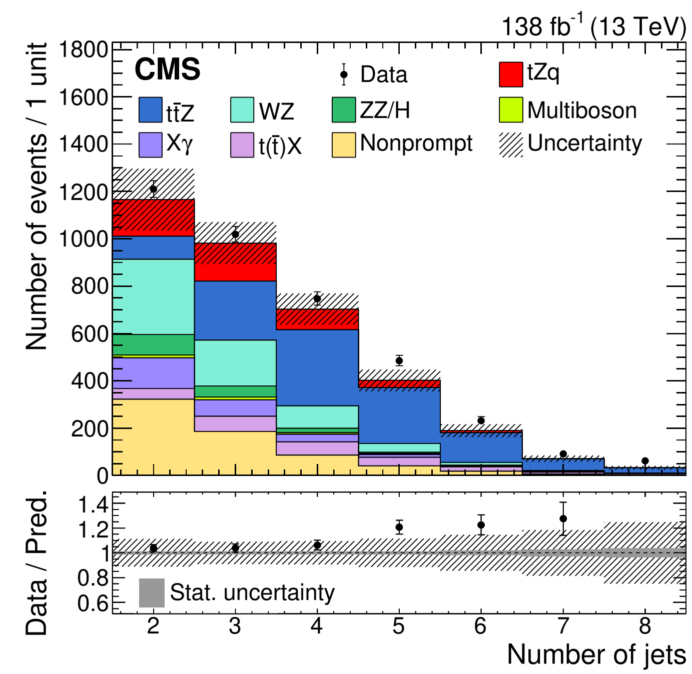
bℓν

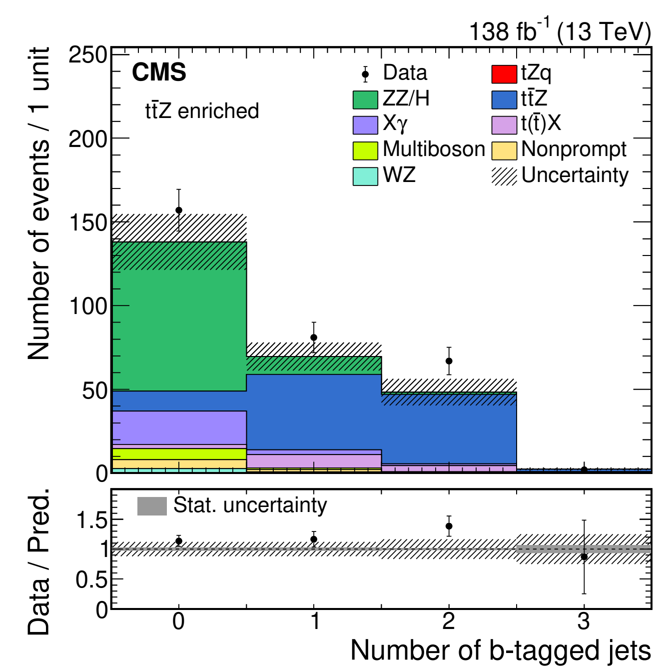
- Selection
  - 3 leptons, Z candidate ≥2 jets, ≥1 b-jet
- 3 regions
  - 1 b-jet, 2-3 jets; ≥4 jets; ≥2 b-jets











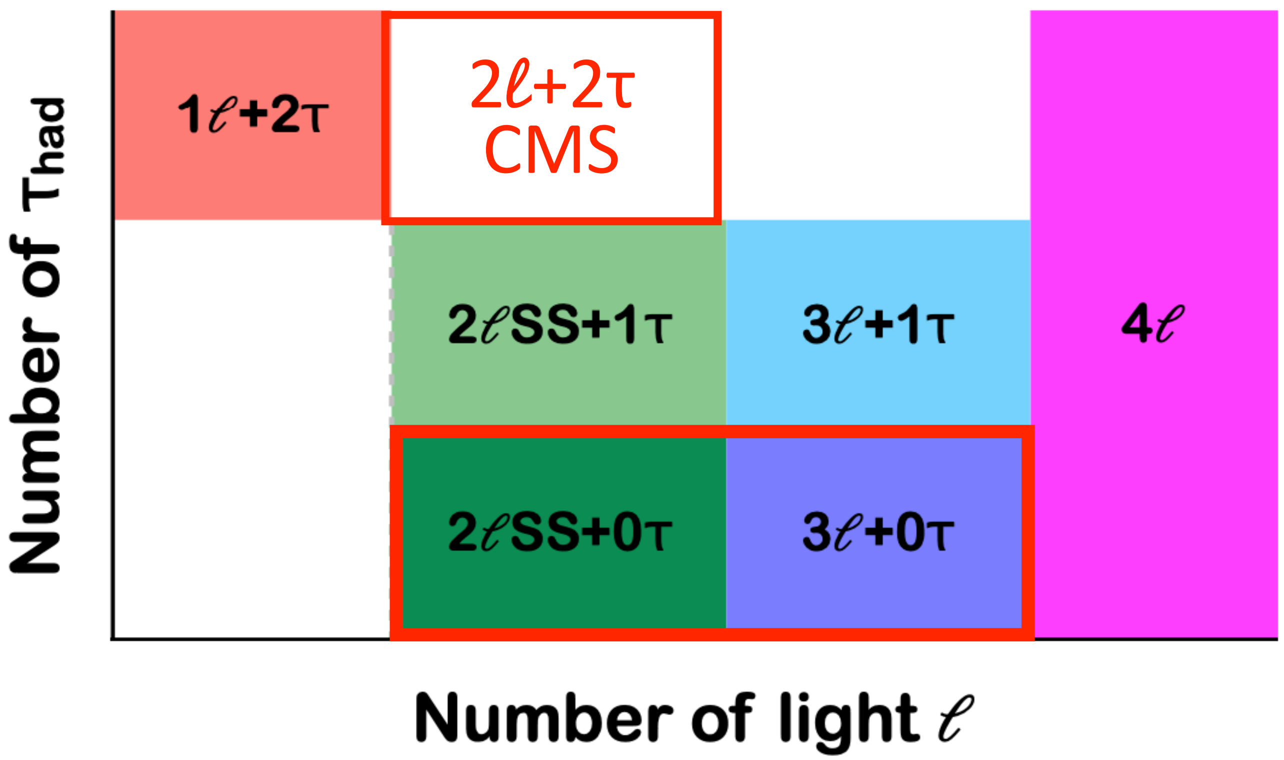
- 1σ pull of ttZ normalisation corresponding to 15% higher cross section

The ttW saga:

Measurements of ttH in multilepton final state

ttH multilepton - ATLAS

ATLAS-CONF-2019-045



- Most sensitive channels
  - 2LSS and 3L
- Irreducible background
  - **$t\bar{t}$Z/W**, VV
- Reducible background
  - non-prompt leptons, charge misID, photon conversions

Basic selection (simplified)

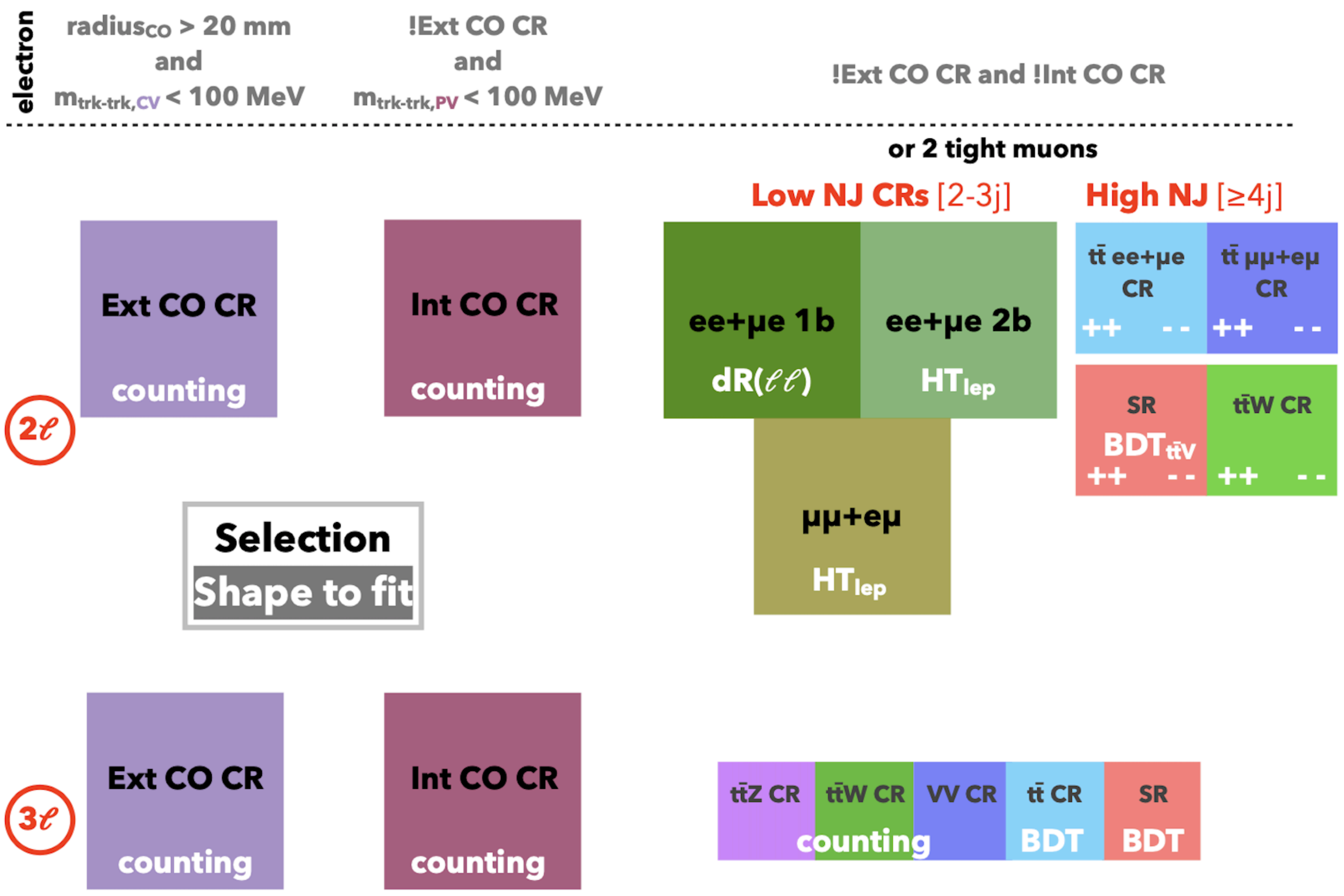
2LSS	3L
$N_{jets} \geq 4$, $N_b \geq 1$	$N_{jets} \geq 2$, $N_b \geq 1$ $\lvert m_{\ell\ell} - m_Z \rvert > 10$ GeV

- MVA discriminants are trained to separate ttH from ttV and tt+jets
- Lower jet multiplicity regions are used to control backgrounds

- Fake background estimate motivates CR definition
- New MC-based template method with 4 free parameters in the fit for various components:
  - HF electron, HF muon
  - material conversions
  - internal conversions

Signal extraction

- BDTs are trained to separate ttH from ttV and ttH from tt+jets (i.e. non-prompt background)



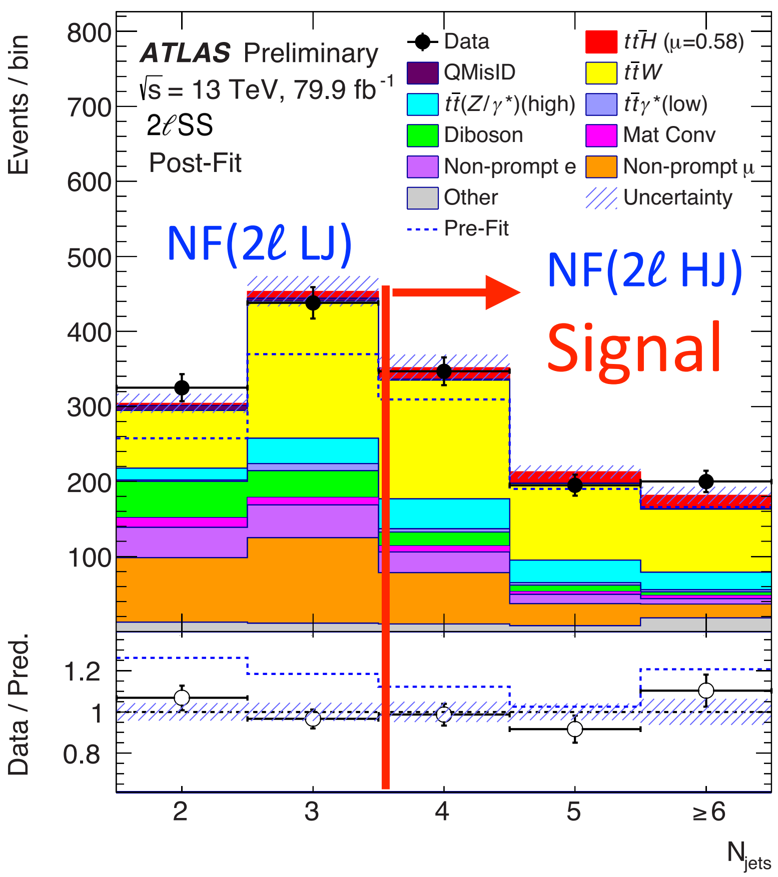
In 2LSS and 3L 17 subcategories based on lepton flavour, charge and b-tagging to control lepton non-prompt, conversions and ttW, ttZ and VV

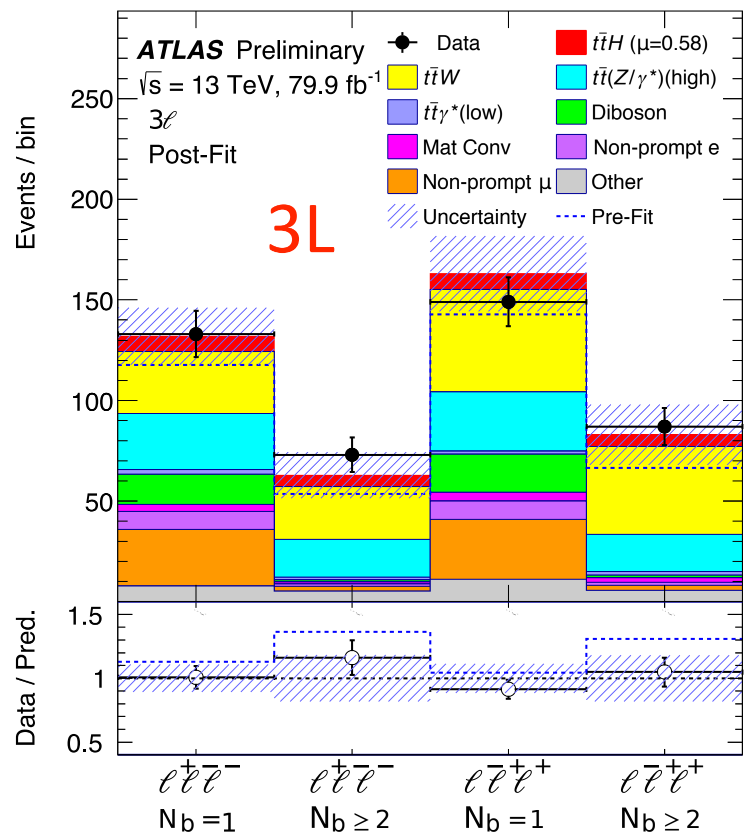
3 signal regions

Control regions

ATLAS-CONF-2019-045

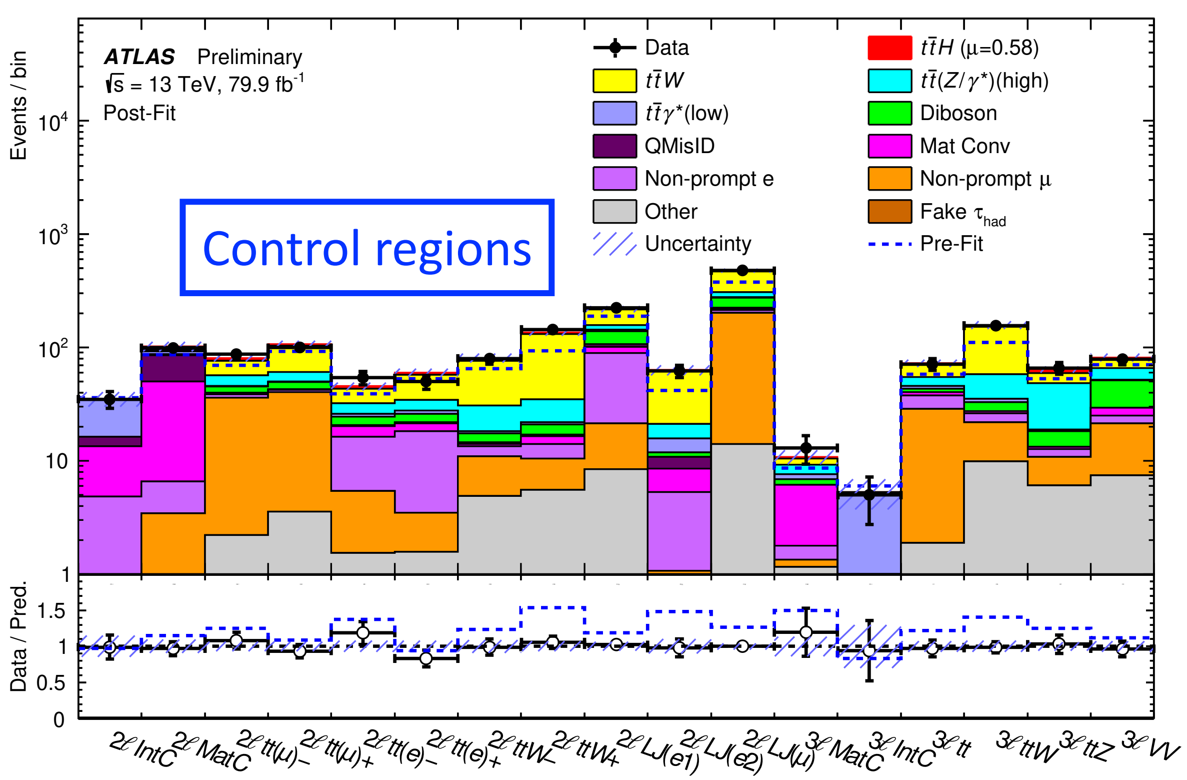
2LSS





Plots after basic selection including low jet multiplicity

Prefit distributions show tension between data and prediction



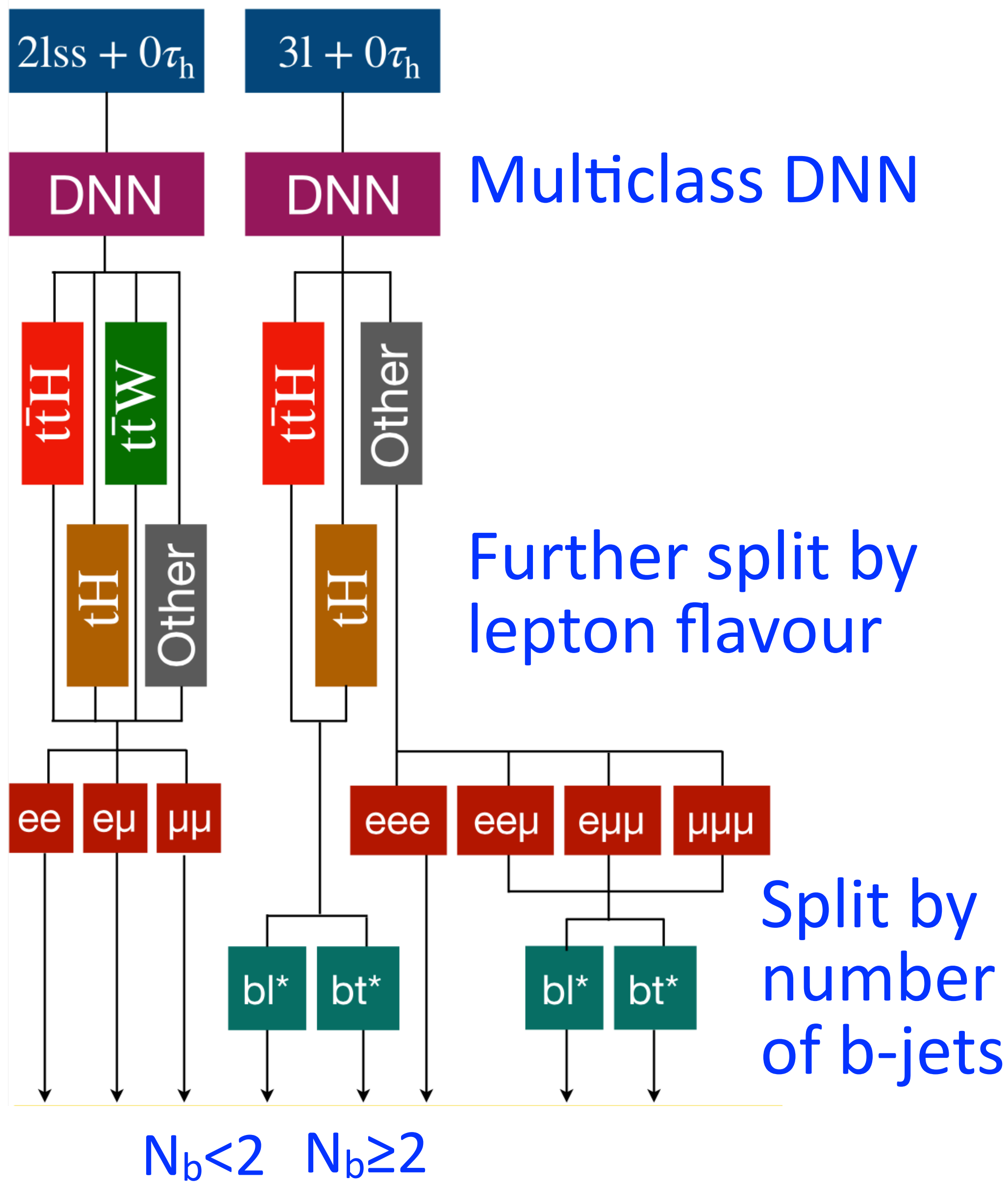
Fit setup	Baseline	Alternativ
μ_{ttH}	$0.58^{+0.36}_{-0.33}$	$0.70^{+0.36}_{-0.33}$
NF (2L LJ)	$1.56^{+0.30}_{-0.28}$	$1.39^{+0.17}_{-0.16}$
NF (2L HJ)	$1.26^{+0.19}_{-0.18}$	
NF (3L)	$1.68^{+0.30}_{-0.28}$	

3 free parameters for ttW to minimise impact of data/MC tension in CRs

$1.67^{+0.20}_{-0.19}$ wrt to YR4

ttH multilepton - CMS

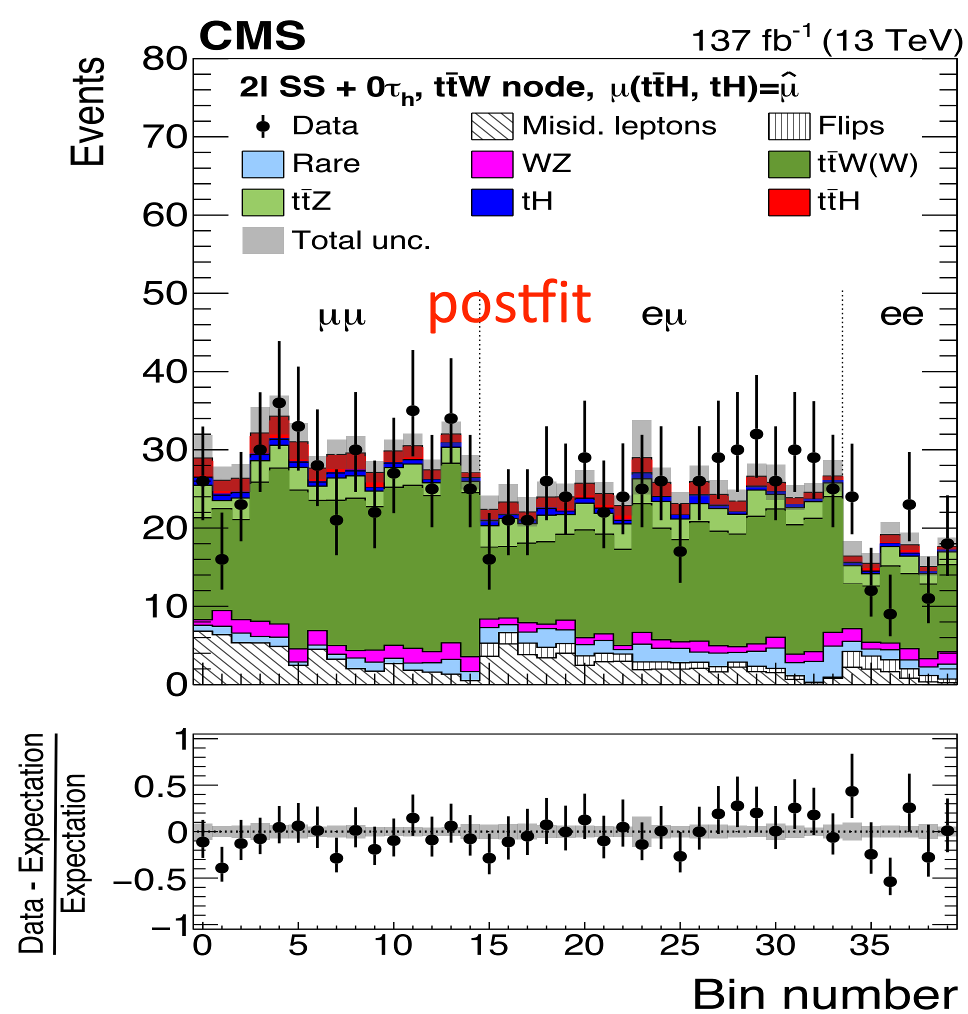
- Simultaneous measurement of ttH and tH
- Channels with e/μ/hadronic τ are considered
- Concentrate on 2LSS and 3L



Backgrounds

Eur. Phys. J. C 81 (2021) 378

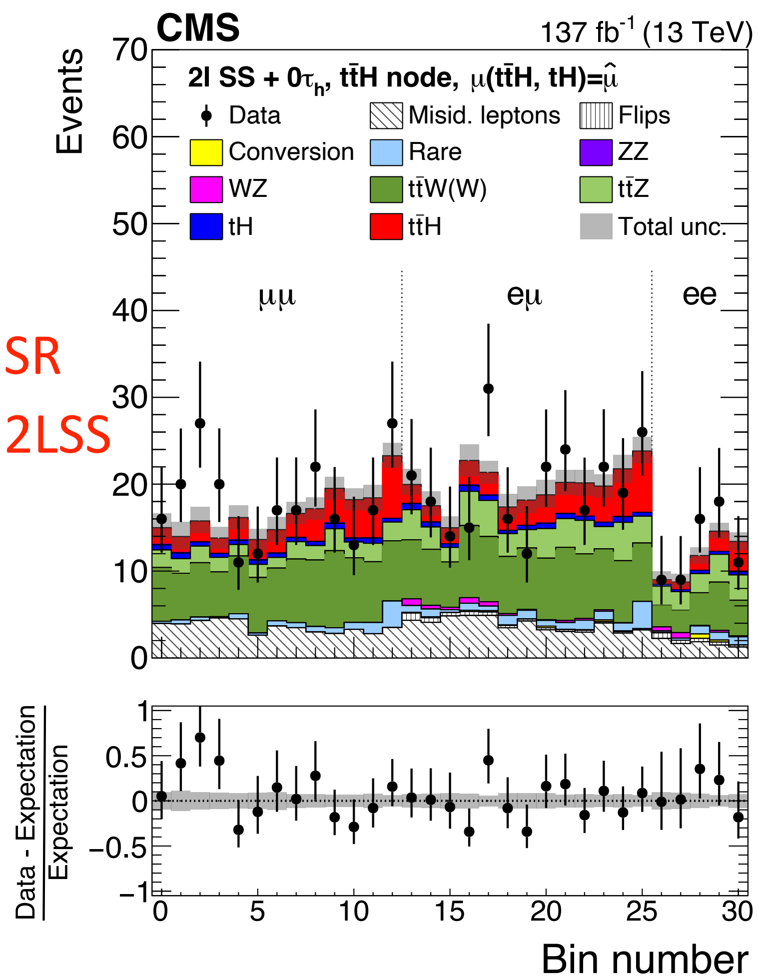
- Non-prompt leptons
  - DD "misidentification probability" method (similar to Fake factor)
- ttZ/WZ CR:
  - Z candidate, no b-jet requirement
  - split in N_{jets} and N_b to separate ttZ and WZ



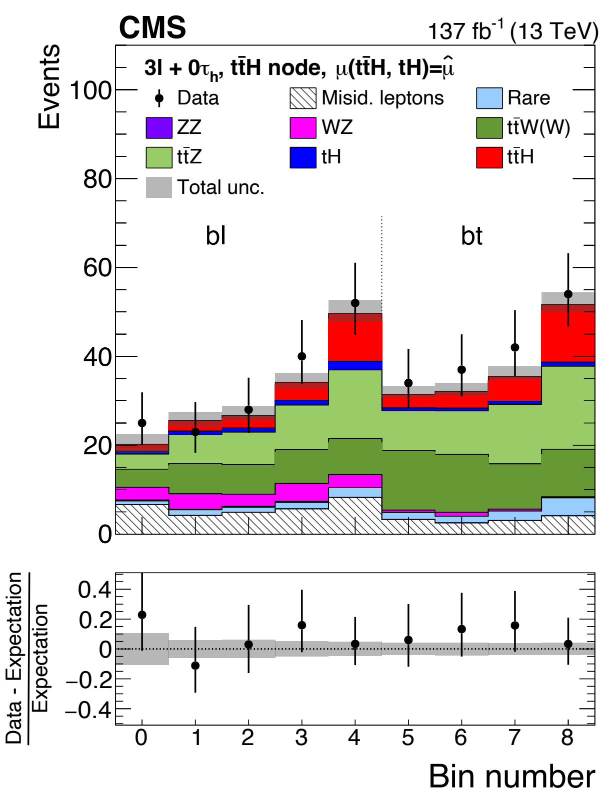
ttW CR defined by DNN

Results

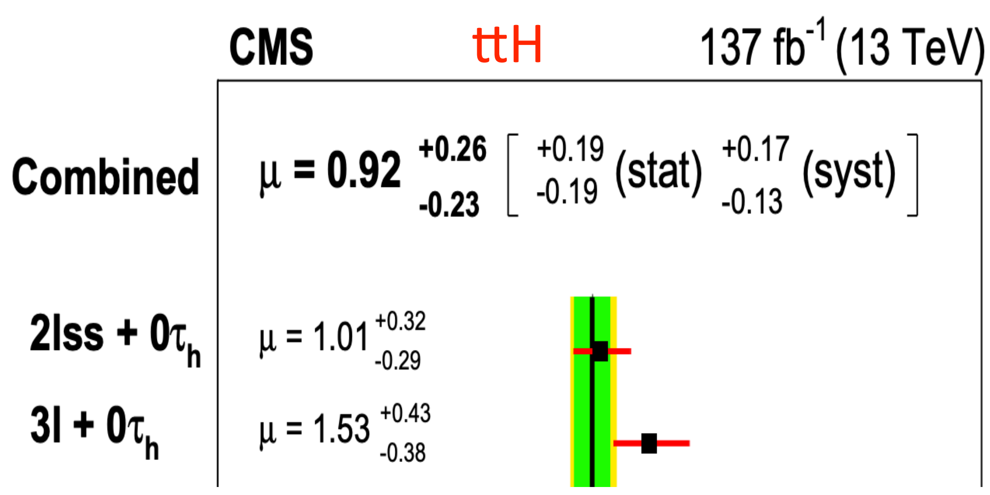
SR 2LSS

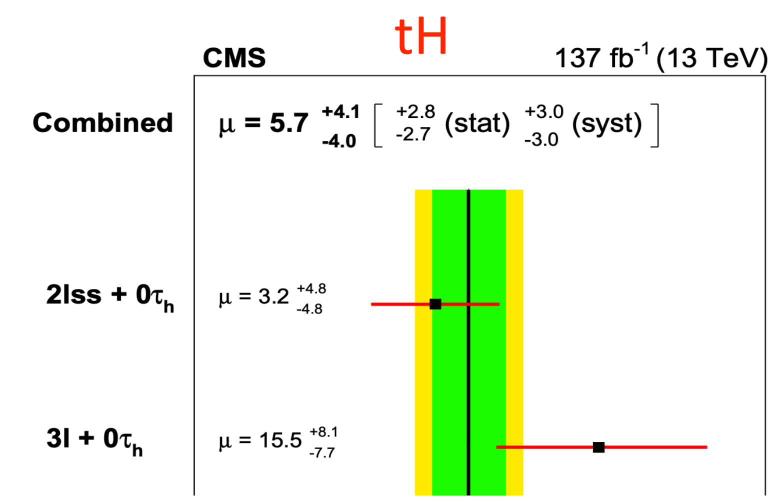


SR 3L



Free parameters of the fit: μ_{ttH}, μ_{tH}, μ_{ttW}, μ_{ttZ}

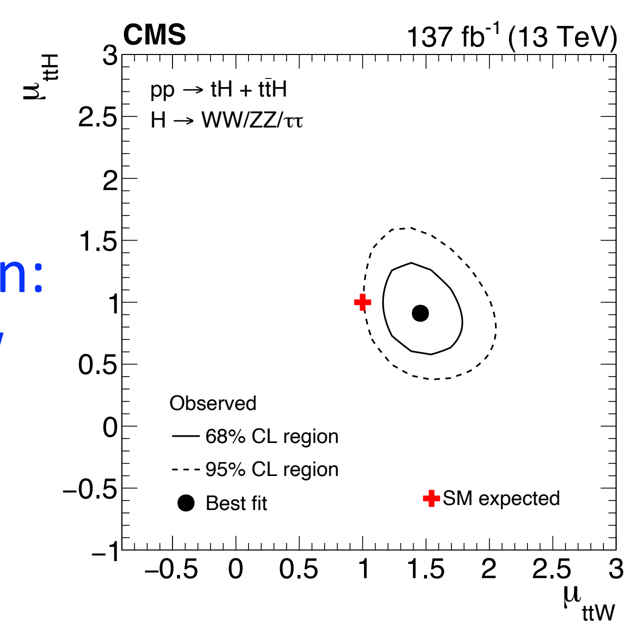


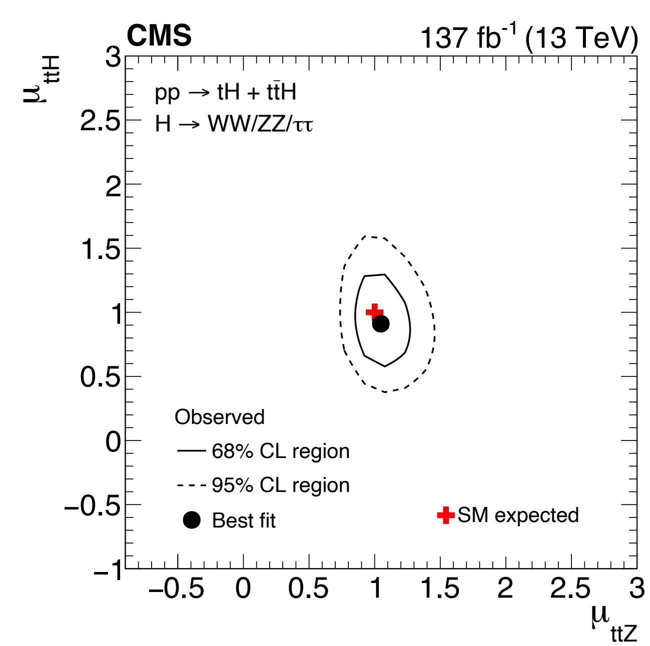


μ_{ttW} = 1.43 ± 0.21 (stat+sys), σ_{ttW} = 650 fb

μ_{ttZ} = 1.03 ± 0.14 (stat+sys), σ_{ttZ} = 839 fb

Correlation: μ_{ttH} vs μ_{ttW}



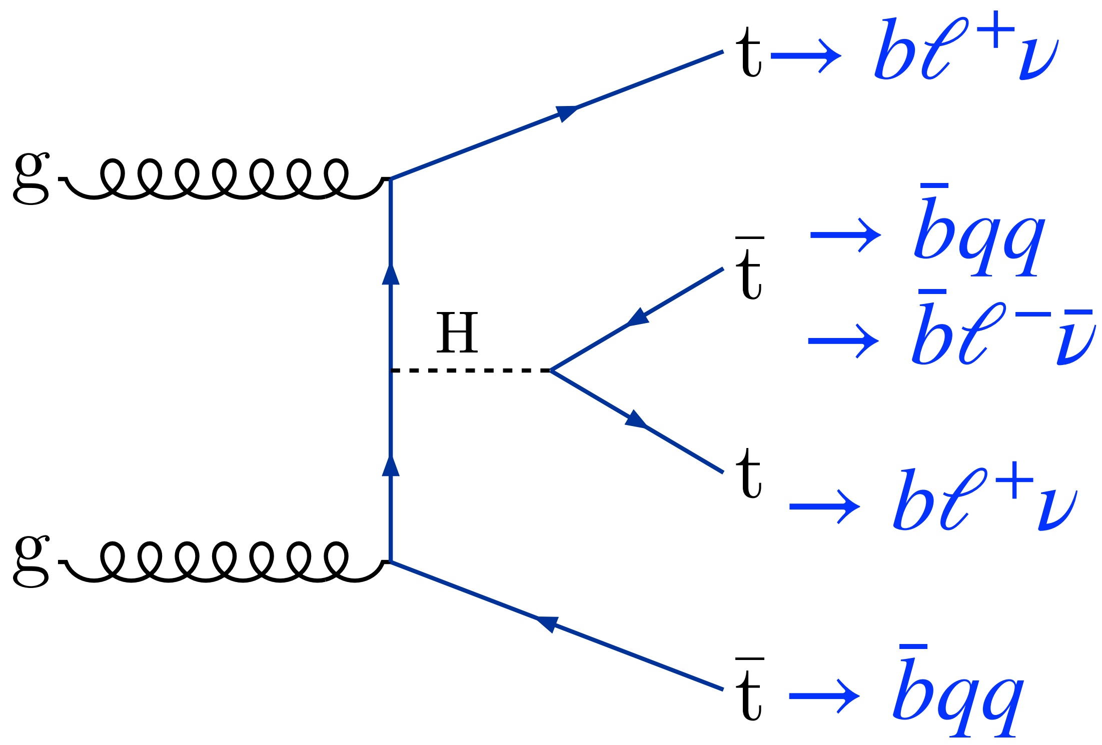


Correlation: μ_{ttW} vs μ_{ttZ}

Searches for 4-top quark production in multilepton final state

SM 4-top search CMS

Eur. Phys. J. C 80 (2020) 75



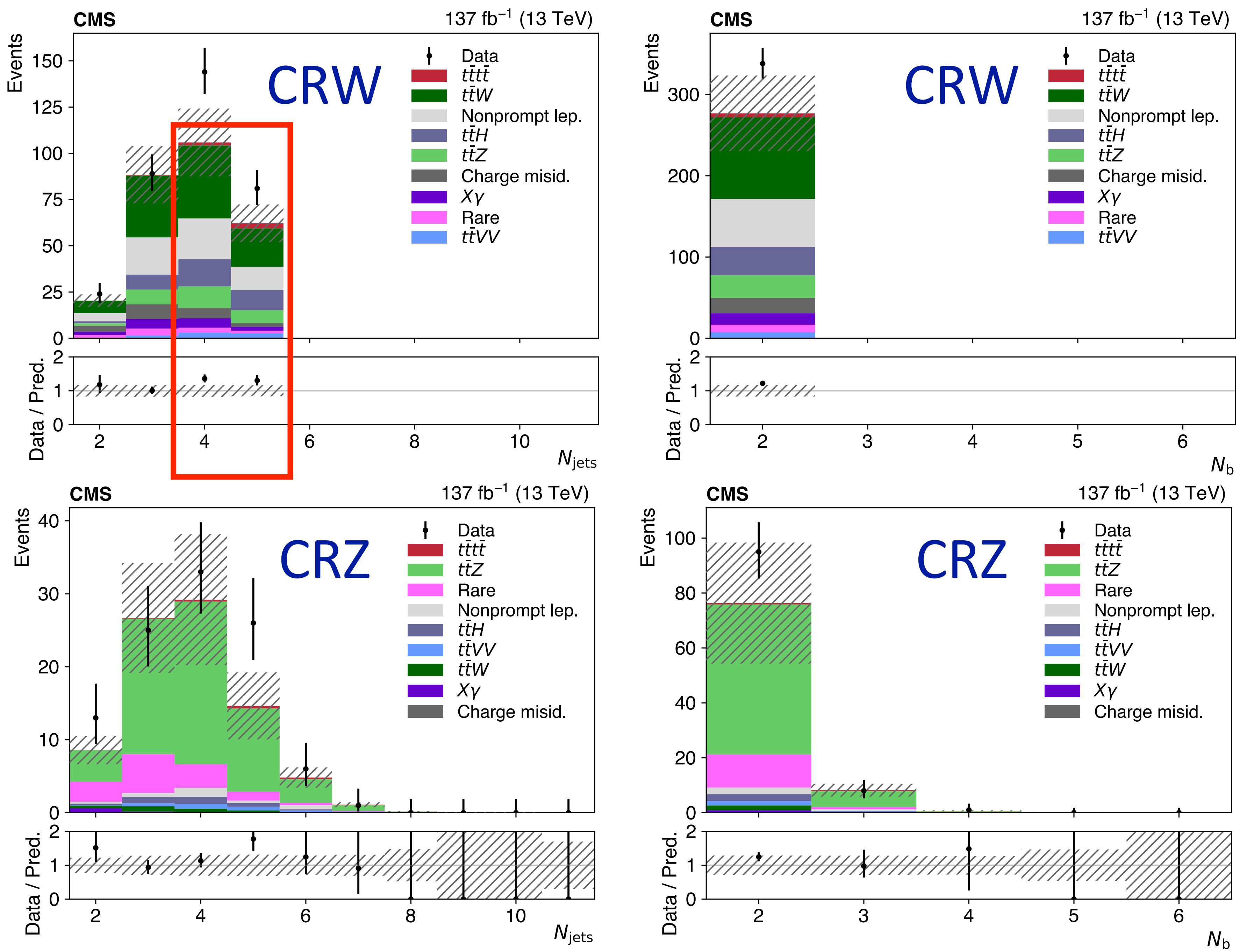
- 2ℓSS and 3ℓ channel signature
  - 4 b-jets from 4 top quarks
  - 4 or 2 jets from W
- Baseline selection
  - ≥2 jets, ≥2 b-jets
  - H_T>300 GeV
  - E_T^{miss} > 50 GeV
  - 30 GeV $Z(\ell^{+}\ell^{-})$ mass window cut

N_ℓ	N_b	N_{jets}	Region
		≤ 5	CRW
		6	SR1
	2	7	SR2
		≥ 8	SR3
2		5	SR4
		6	SR5
	3	7	SR6
		≥ 8	SR7
	≥ 4	≥ 5	SR8
		5	SR9
	2	6	SR10
≥ 3		≥ 7	SR11
		4	SR12
	≥ 3	5	SR13
		≥ 6	SR14
Inverted resonance veto			CRZ

BDT option of analysis does not have ttW CR

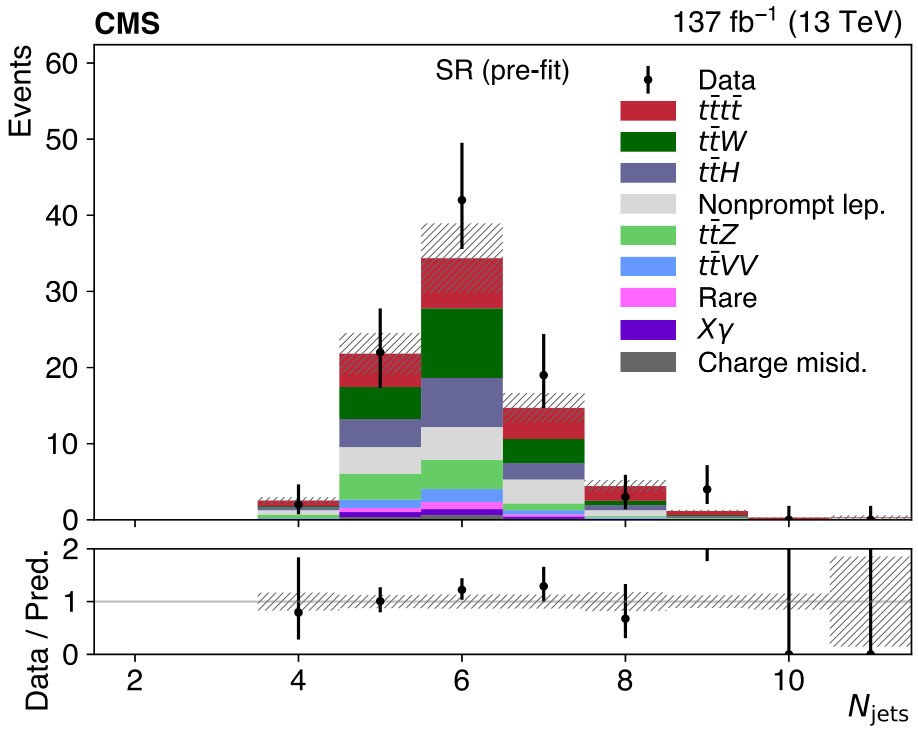
4-top search: ttW/ttZ CR

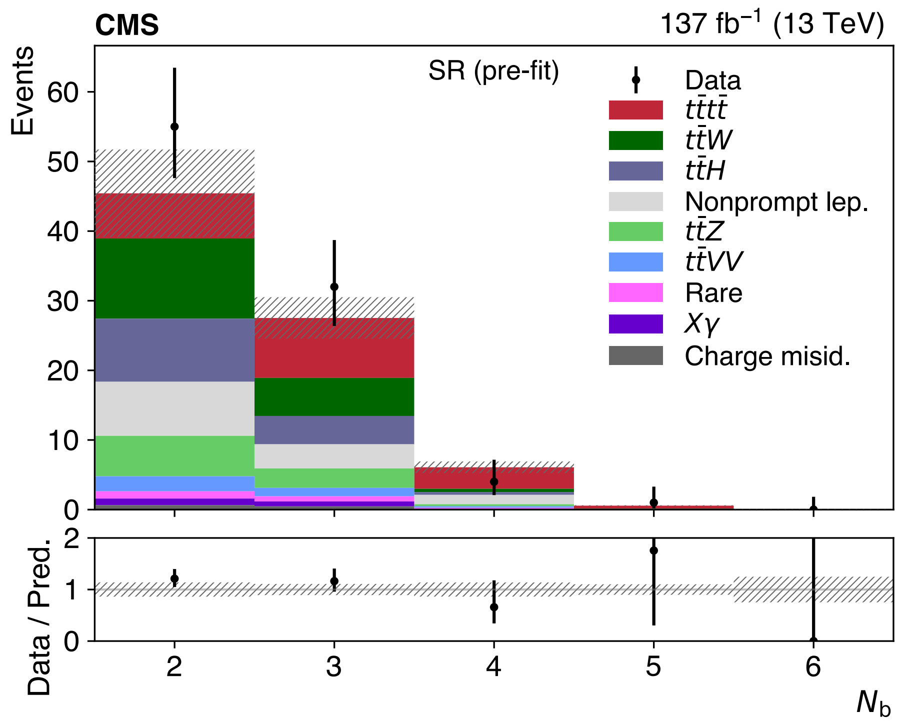
Prefit plots



- DD methods for non-prompt and charge misID leptons
- ttW CR: excess of data over prediction for nj=4,5
- ttZ CR: excess in nj=5
- ttW/ttZ normalisation in the fit: 40% Gaussian prior

ttW/ttZ corrections





prefit plots in SR after ttV correction

Corrections of ttW, ttZ simulation (MG5_aMC@NLO+Pythia8)

with extra jets

- compare light-flavour jet multiplicity in dilepton $t\bar{t}$ events in data and MG5_aMC@NLO+Pythia8
- derive correction and apply as weight to ttV
- weights vary between 1.46 and 0.77 for 1 to 4 additional jets

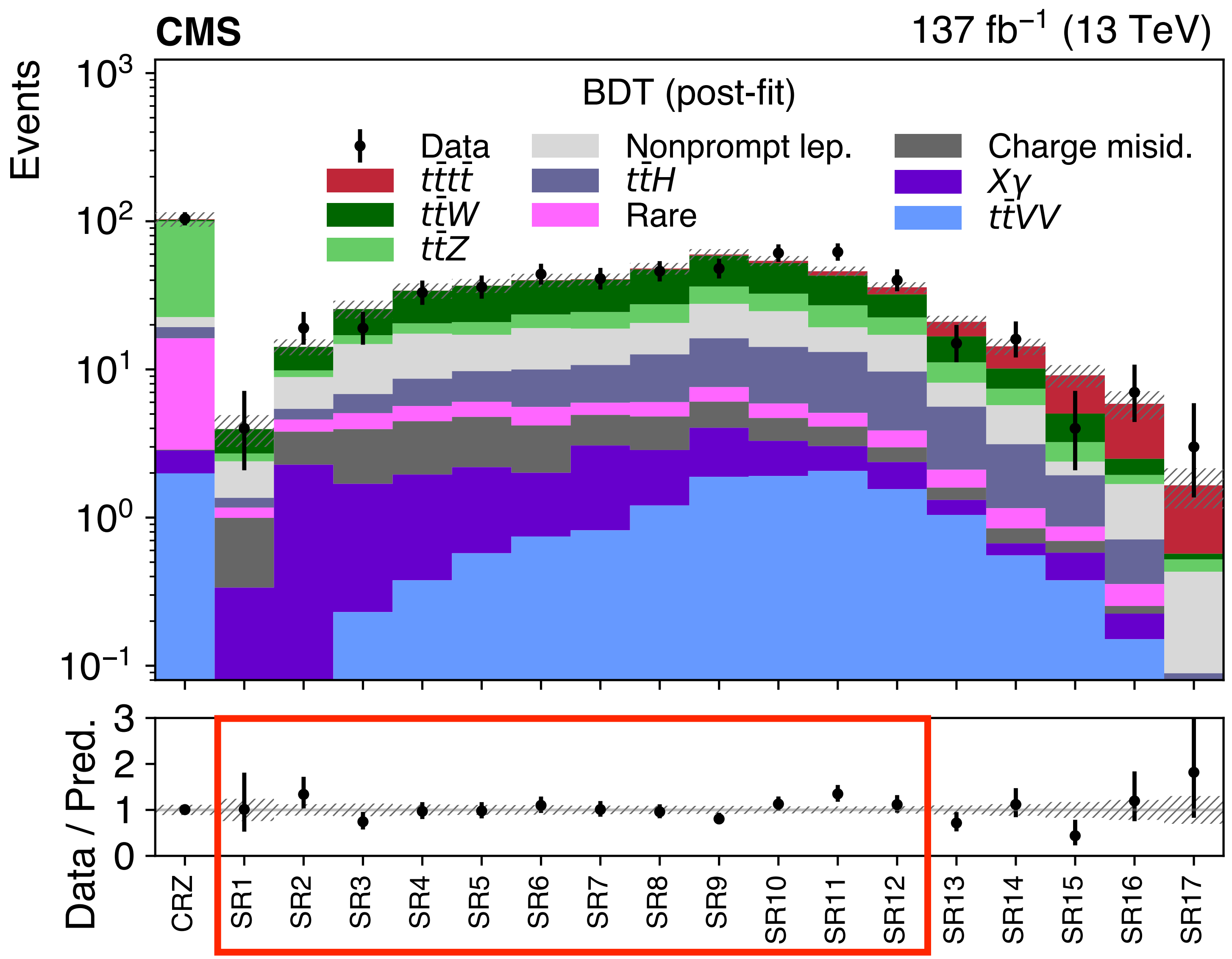
with extra b-jets

- factor of 1.7±0.6 is applied to improve modelling of extra heavy flavour jets
- value is based on measured ratio of $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ ratio
- 70% increase of events with additional $b\bar{b}$ pair

Uncertainties on corrections are included in systematics

4-top search result

Post-fit $t\bar{t}W/t\bar{t}Z$ scaled by 1.3±0.2



- Dominant systematics:
  - modelling of additional b-jets (11%)
  - JES (9%) JER (6%)
  - b-tagging (6%)

- In BDT analysis ttW is constrained by regions with low BDT score

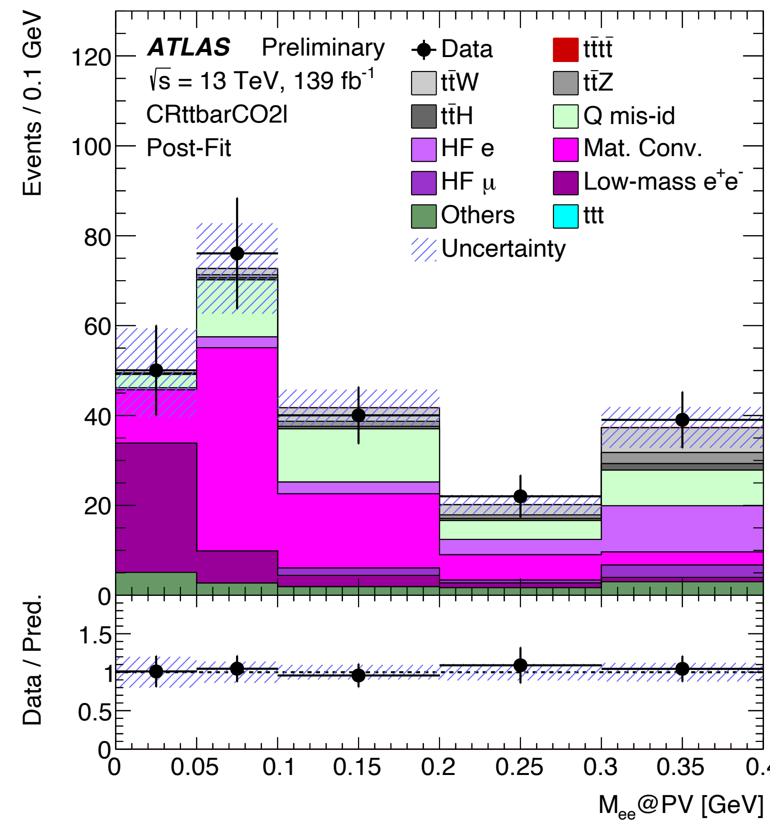
$$\sigma(t\bar{t}t\bar{t}) = 12.6^{+5.8}_{-5.2} \text{ fb}$$

2.6 (2.7) σ observed (expected) significance

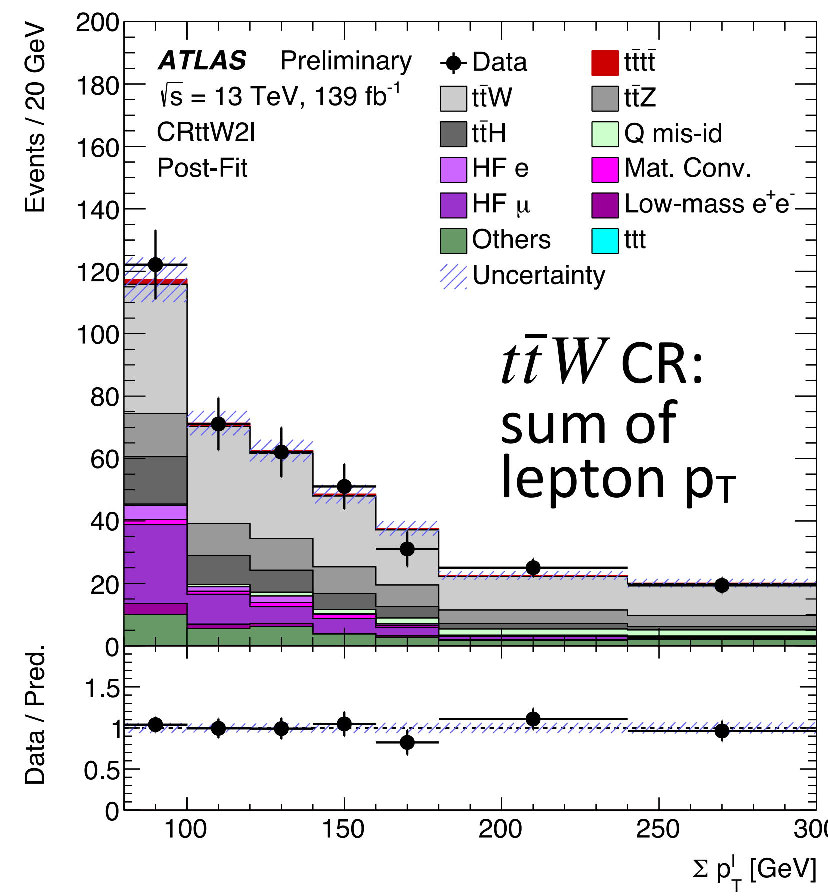
4-top: ATLAS

Eur. Phys. J. C 80 (2020) 1085   139 fb^{-1}

- Basic selection
  - $\geq$4 jets, $\geq$1 b-jets
  - $|m_{\ell\ell} - m_Z| > 10$ GeV for OS SF pairs in 3L selection
- Fake background estimate motivates CR definition
- MC-based template method with 4 free parameters included in the signal extraction fit
  - HF electron, HF muon
  - material conversions
  - internal conversions
- $t\bar{t}W$ background normalisation is free in the fit

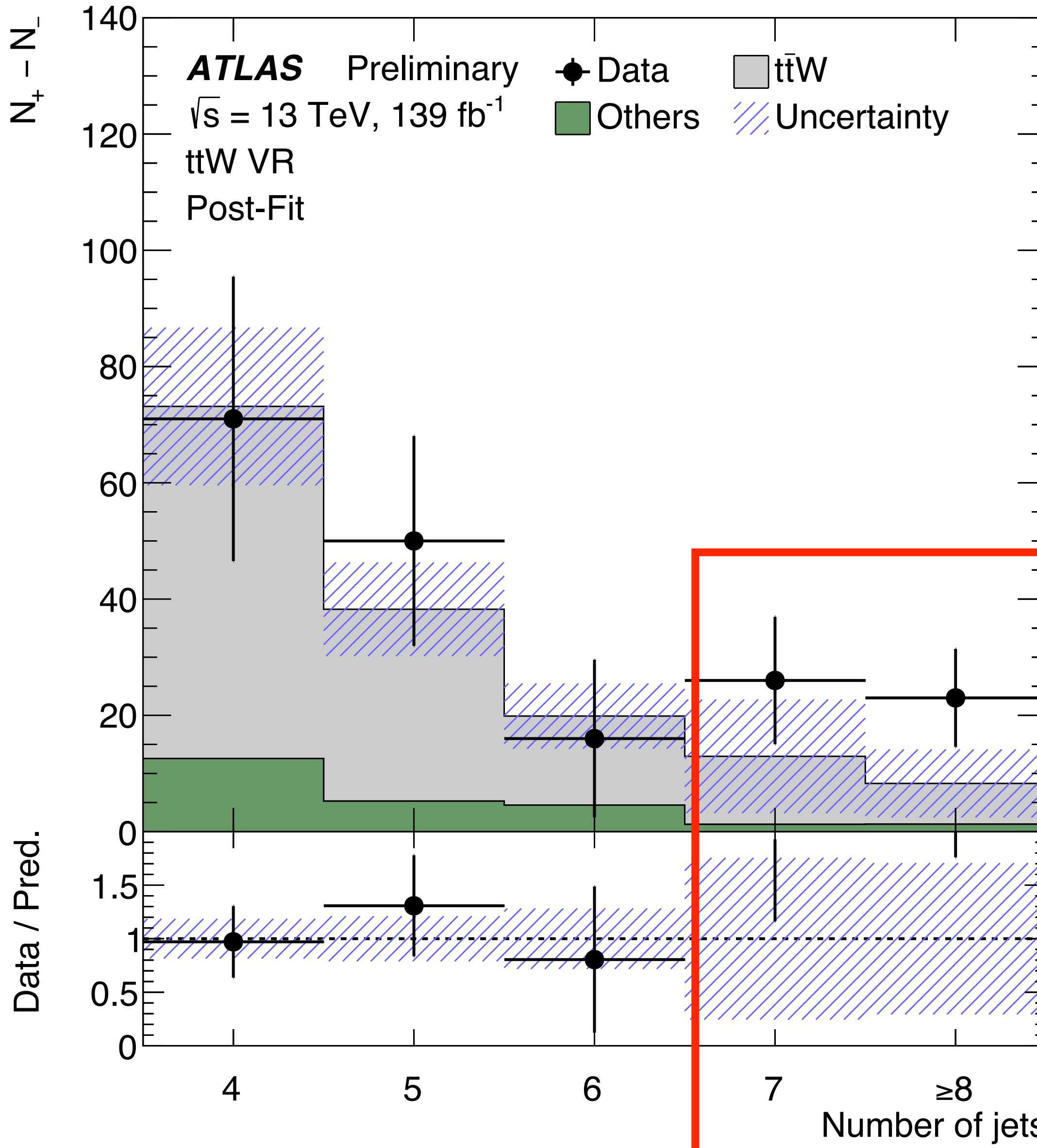


Material conversion CR



ttX background model

- $t\bar{t}W$ validation region:
  - ≥4 jets, ≥2 b-jets
  - plot $N_+ - N_-$ to suppress all charge symmetric backgrounds
- Systematic uncertainties on $t\bar{t}W$:
  - 125% on $t\bar{t}W$+7 jets
  - 300% on $t\bar{t}W$+8 jets
  - 50% on $t\bar{t}W$+3b, $t\bar{t}W$≥4b
- Additional 50% uncertainty is applied to $t\bar{t}Z$ and $t\bar{t}H$ with 3 and ≥4b jets
- 100% uncertainty on 3-top cross section and additional 50% on ttt+b

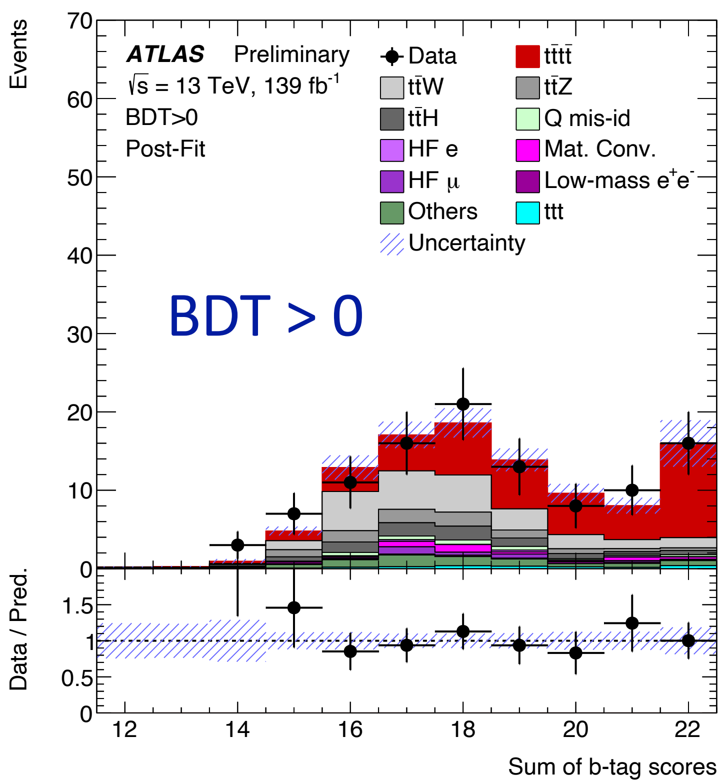


Parameter	$NF_{t\bar{t}W}$	NF_{CO}	NF_{γ^*}	NF_{HFe}	$NF_{HF\mu}$
Value	1.6 ± 0.3	1.6 ± 0.5	0.9 ± 0.4	0.8 ± 0.4	1.0 ± 0.4

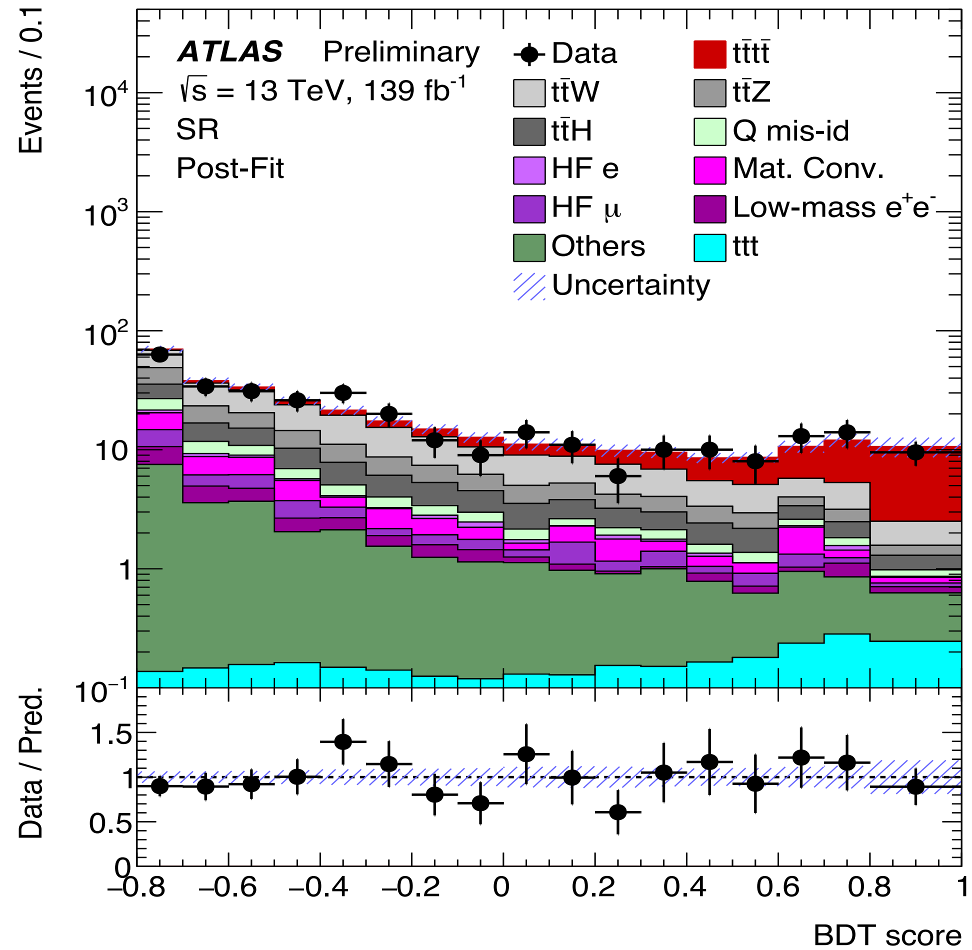
σ_{ttW} = 601 fb

4-top search result

- Inclusive signal region
  - ≥6 jets, ≥2 b-jets, H_T>500 GeV
  - BDT is trained to separate signal from background
- Signal extraction
  - simultaneous fit to BDT score and distributions in 4 CRs



$$\mu = 2.0^{+0.4}_{-0.4}\ \text{(stat)}^{+0.7}_{-0.5}\ \text{(syst) fb}$$



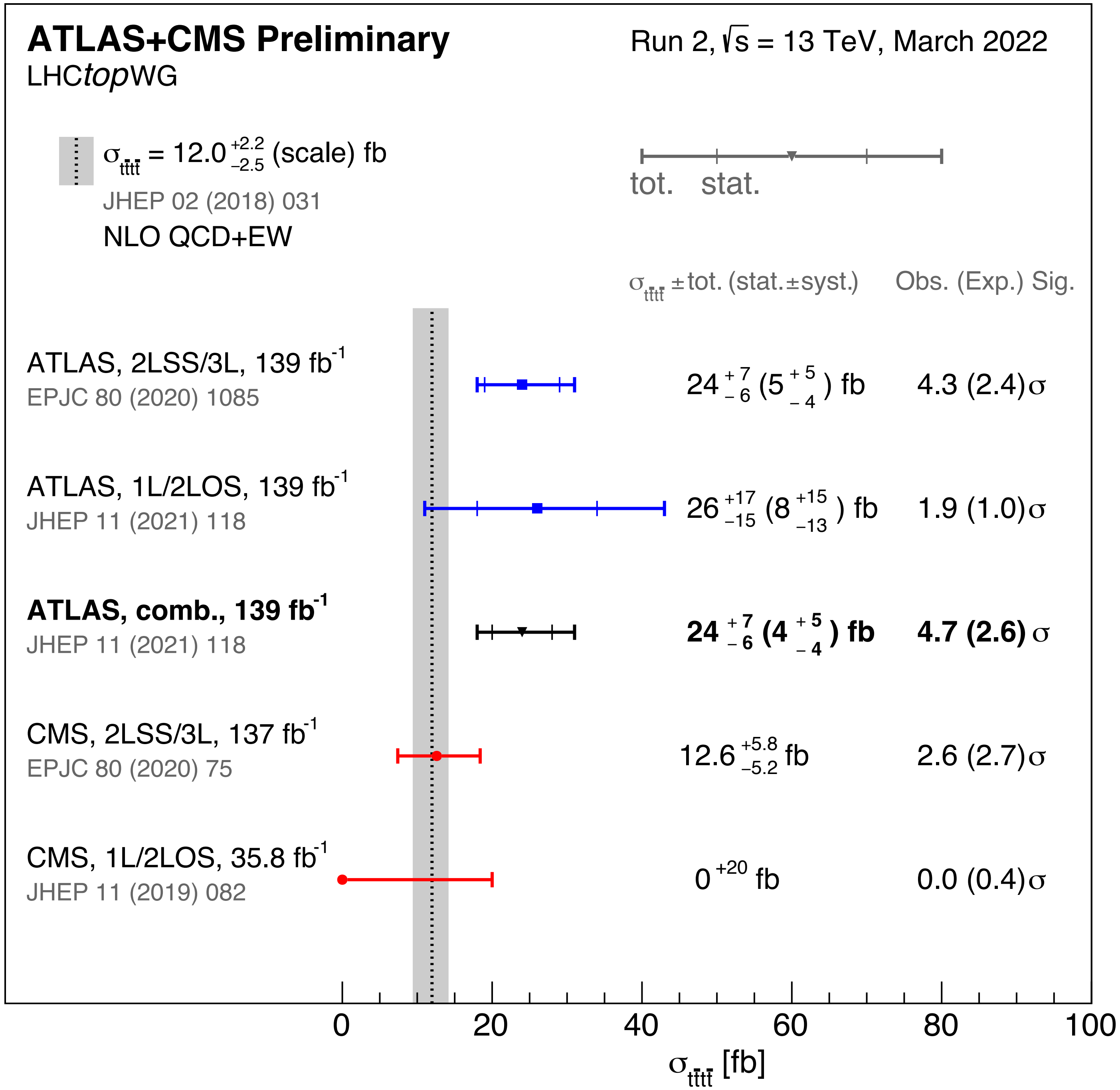
Postfit overall $t\bar{t}W$ background yield in the signal-enriched region with a BDT > 0 increased from 12.4 ± 8.8 events prefit to 23.2 ± 10.1 events, i.e. by a factor of 1.9

4.3 (2.4) σ observed (expected) significance

Evidence for the 4-top quark production

Consistent with SM prediction at 1.7 σ

4-top summary



$t\bar{t}W$ measurement

ttW measurement

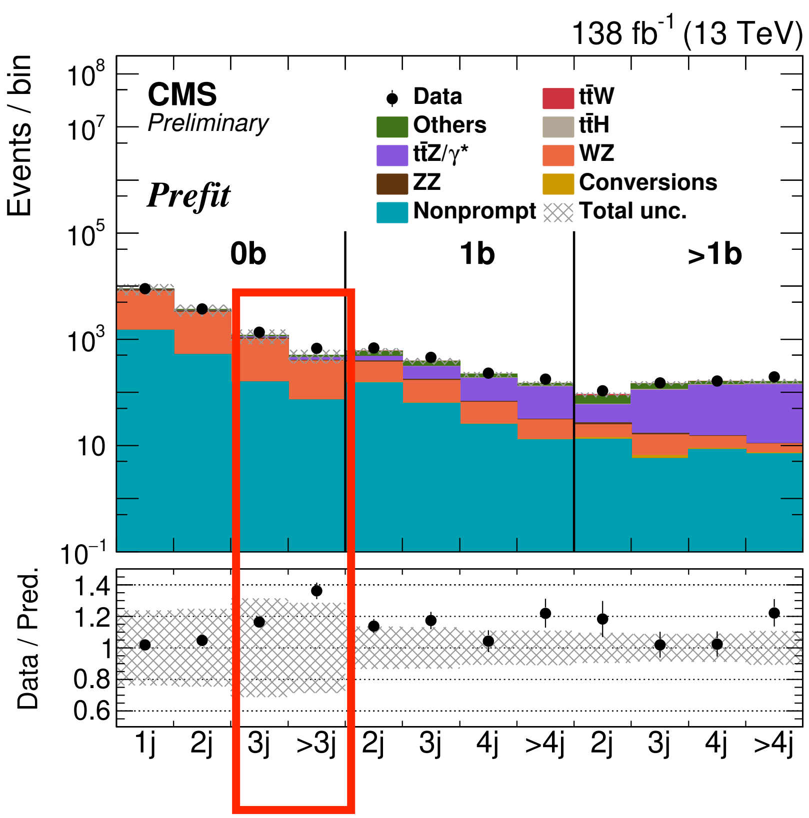
CMS-PAS-TOP-21-01

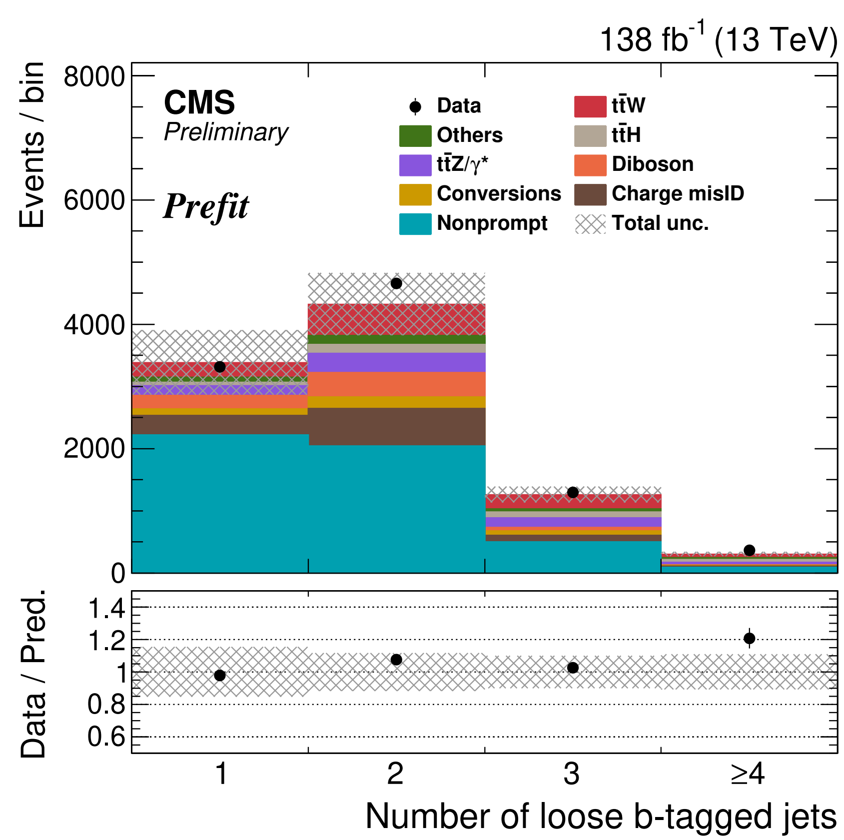
- **Selection**
  - 2 same-sign or 3 leptons
  - Z veto (ee, μμ) in 3L and in (eeSS)
  - $N_{jets} \geq 2$
  - 1b medium or 2b loose in 2LSS
  - at least 1b medium in 3L
  - Large E_T^{mis} in 2LSS

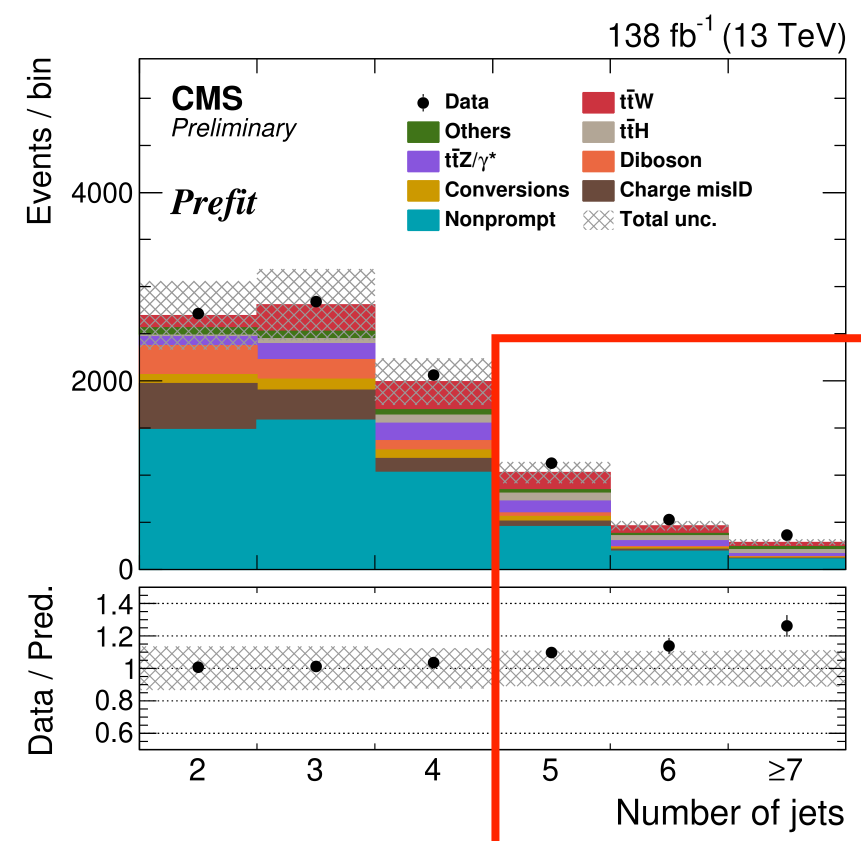
Strategy

- 2LSS: multiclass NN with 4 nodes
  - ttW, non-prompt, ttH/ttZ, ttγ*
- 3L: categorisation based on N_{jets}, N_b, lepton charge
  - fit to m(3L) in each category
- WZ and ttZ CR
- DD non-prompt and charge misID leptons

WZ and ttZ CR

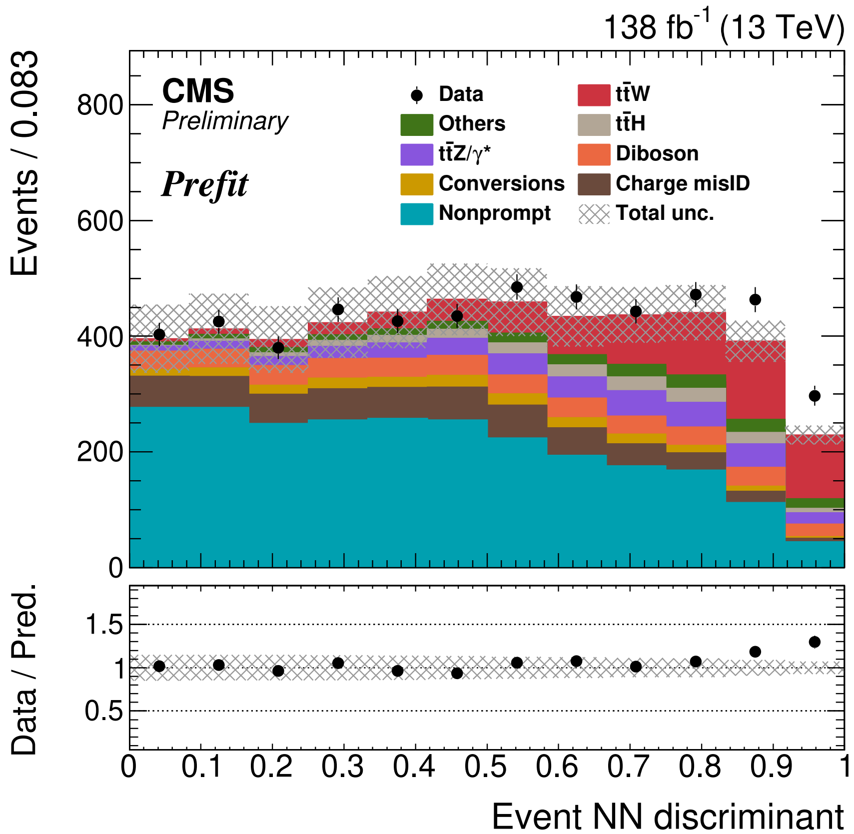


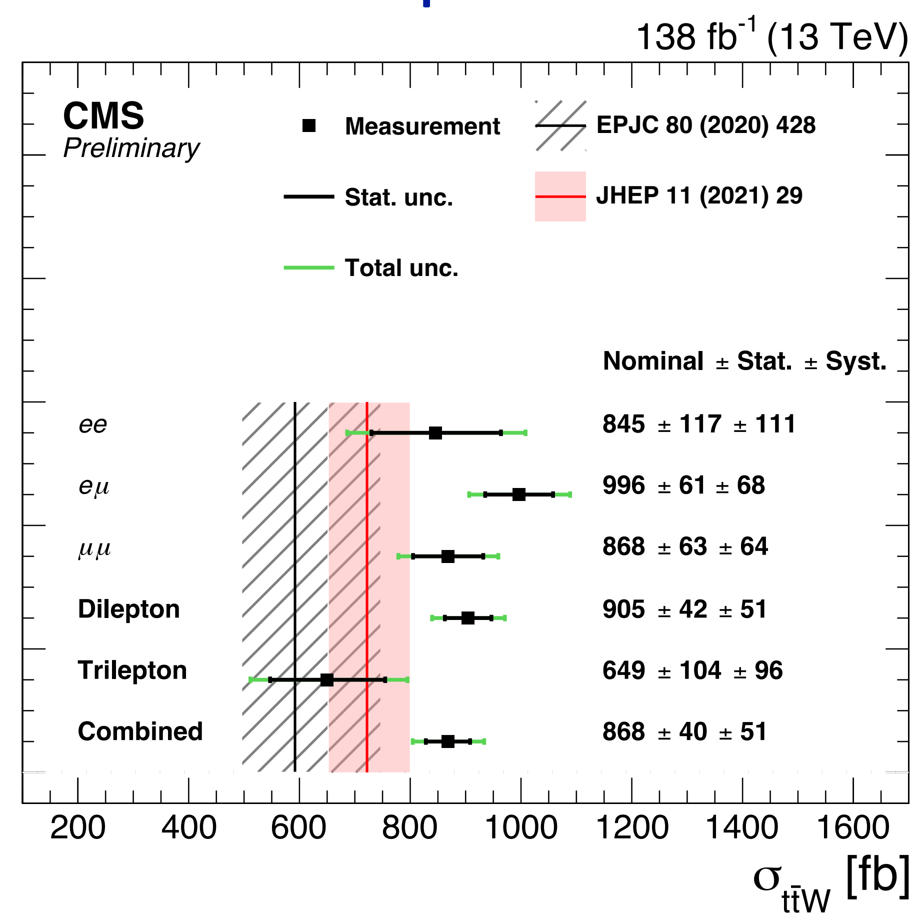




Results

- ttZ, WZ, ZZ normalisations are free parameters of the fit

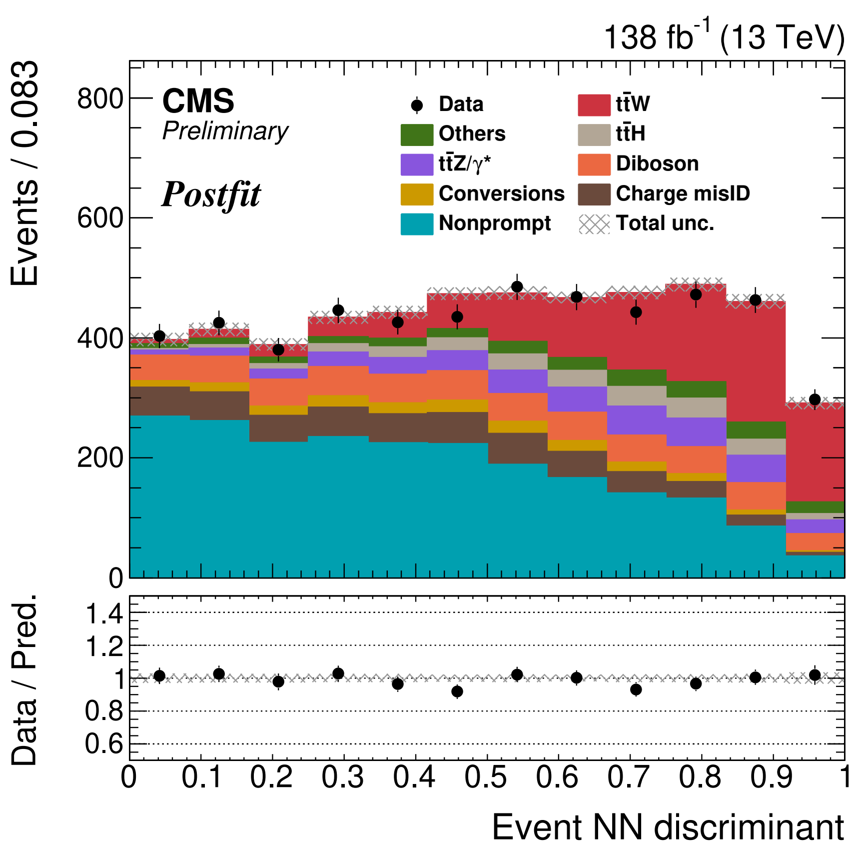


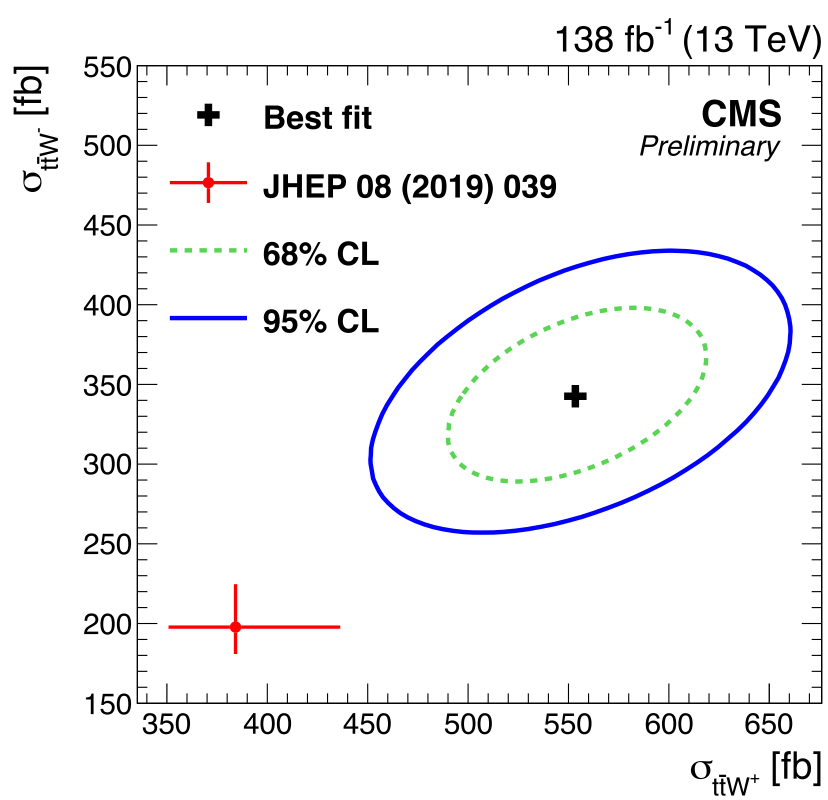


Assumed ttW SM

σ_{ttW} = 592 fb

Measured combined cross section corresponds to μ_{ttW} = 1.47





Significant deviation from prediction for ttW^{+}/ttW^{-} ratio = 1.94+0.37-0.24

What can go wrong in multilepton+b analysis?

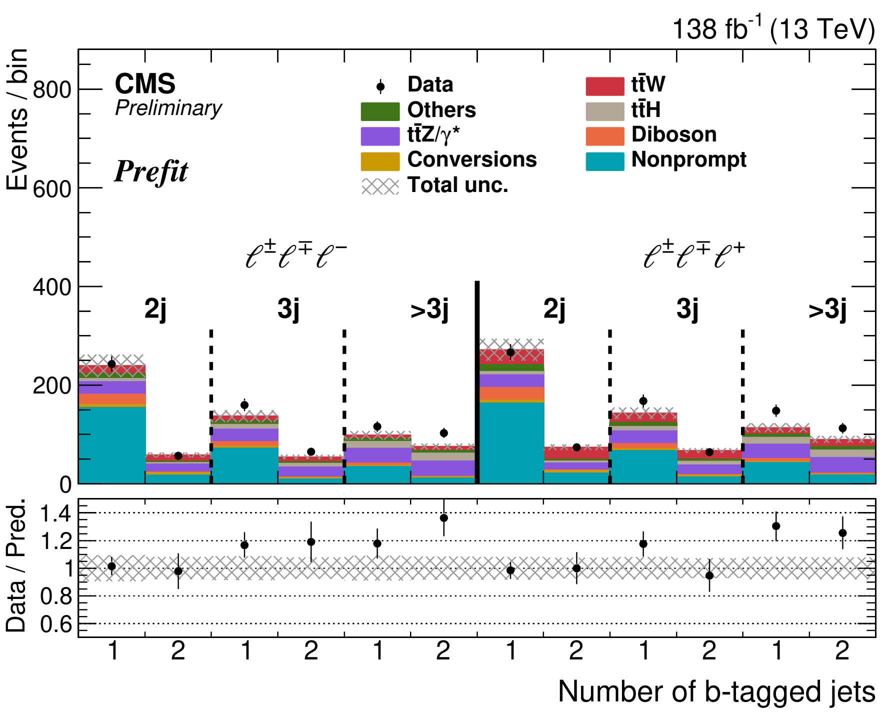
ttV model: normalisation

- ttZ normalisation
  - CMS: NLO with EWK corrections (YR4, on-shell only): 840 fb
  - ATLAS: added off shell contribution: 880 fb

- ttW normalisation
  - NLO in QCD with leading NLO EW corrections (YR4) : 600^{+13}_{-12} fb
  - Current analyses : assumed cross section vary from 592 to 650 fb
  - Can't explain data/prediction tension but it would be nice to use one recommended value in Top and Higgs measurements

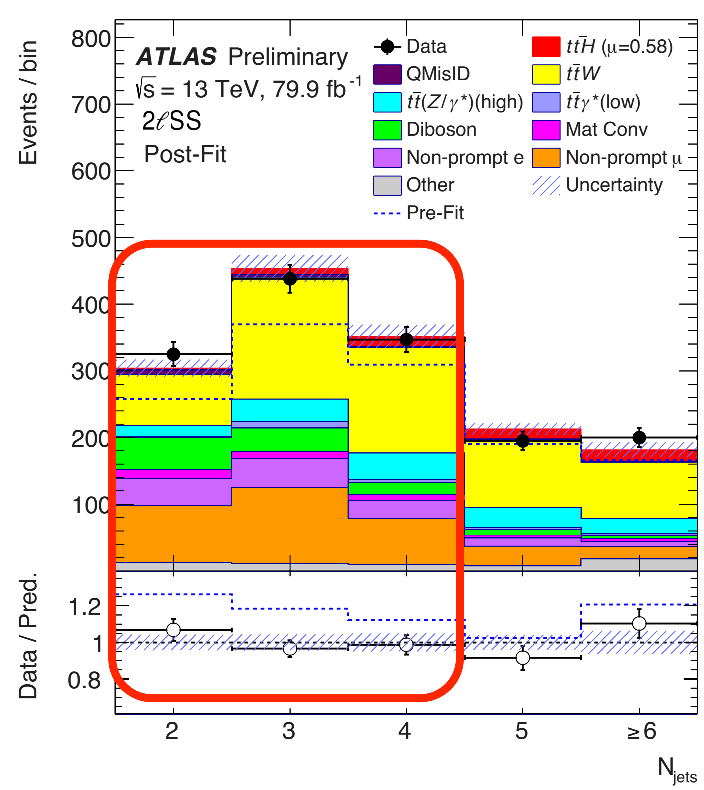
ttW model: MC simulation

- CMS (in ttW measurement):
  - MadGraph5_aMC 2.6.0 at NLO in QCD with $\alpha_s^3\alpha$ term included
  - The $\alpha^3\alpha_s$ term is simulated separately by MadGraph5_aMC
- ATLAS (in 4-top analysis)
  - Sherpa 2.2.1 NLO + 1p@NLO +2p@LO
- There are hints of mismodelling
- Depending on the phase space of the CR fitted μ_{ttW} can be different
  - Can differ between 2LSS and 3L channel

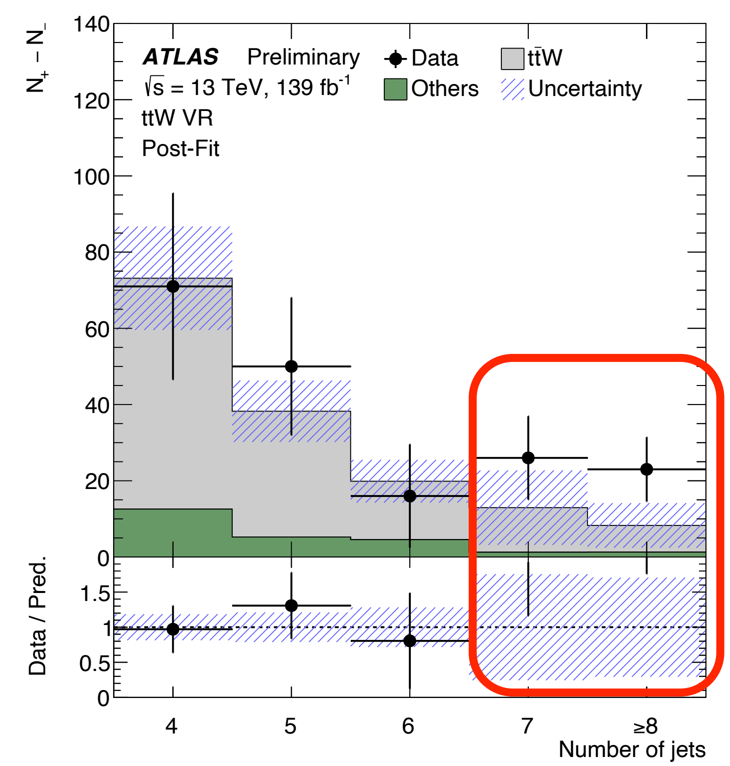


Low purity ttW, hard to make strong conclusion

Low jet multiplicity

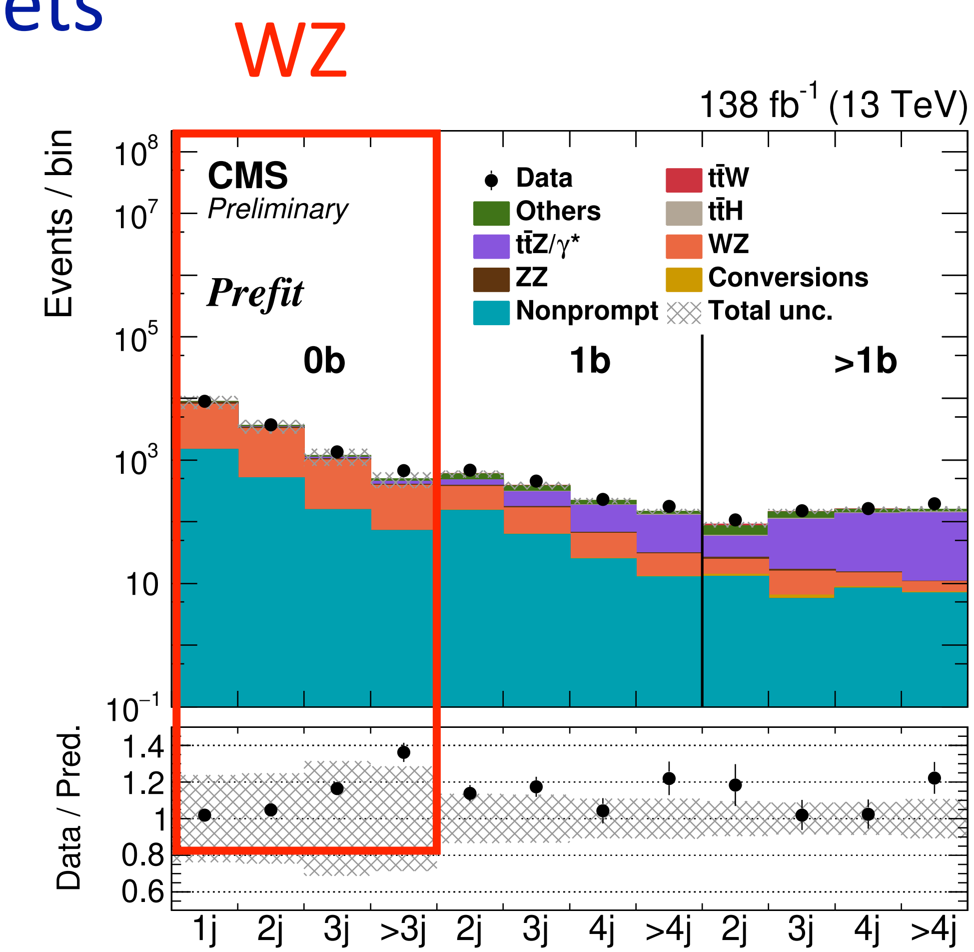


High jet multiplicity



Other MC modelling

- There is sensitivity to
- WZ and ZZ with additional jets and b-jets

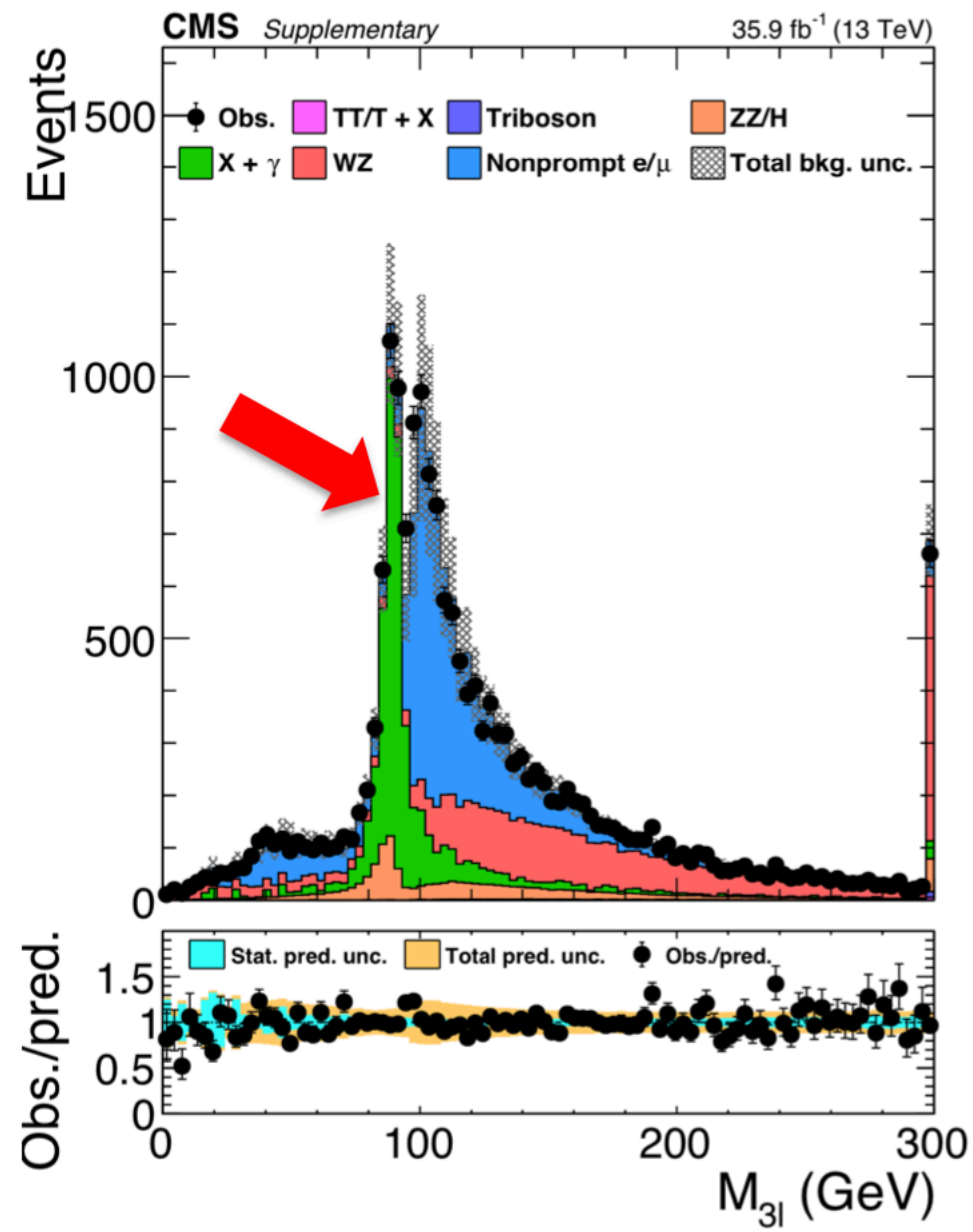


- Modelling of rare backgrounds
  - 3-top
  - ttWW
  - …

- We are sensitive to the tails of distributions and to seemingly small contributions

Fakes: sources

- Fakes: misidentified or non-prompt leptons
- Muons:
  - mainly semileptonic heavy flavor decays (b, c → qW → qeν) plus small contribution from in-flight decays
- Electrons:
  - semileptonic heavy flavor decays (b, c → qW → qeν)
  - photon conversions γ->e+e-
  - light flavor jets: hadrons misidentified as electrons
- In 2LSS and 3L mainly from tt and DY production
- Photon conversions are taken from MC simulation by CMS since they are very well modelled

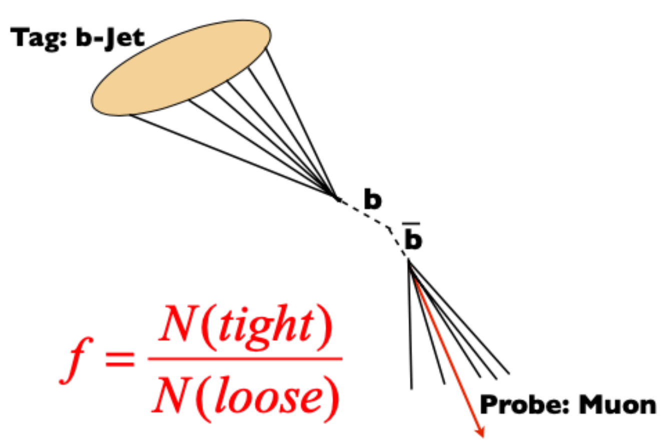


Fakes: methods

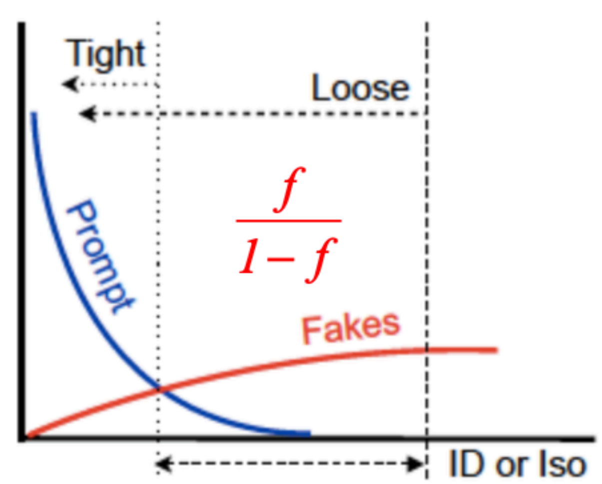
- Data-driven "tight-to-loose" (CMS)
- Matrix Method in ATLAS is similar
  - Require to evaluate TL probability (*f*)
  - CMS region to measure f is far away from signal region —> fake composition and kinematics might be different
  - ATLAS measured f very close to signal region —> potential signal contamination

- Recent ATLAS measurements use template method
  - Relies significantly on MC
  - Obtain MC templates for the different fakes sources
  - Perform a fit to data in dedicated CRs to extract normalisation factors

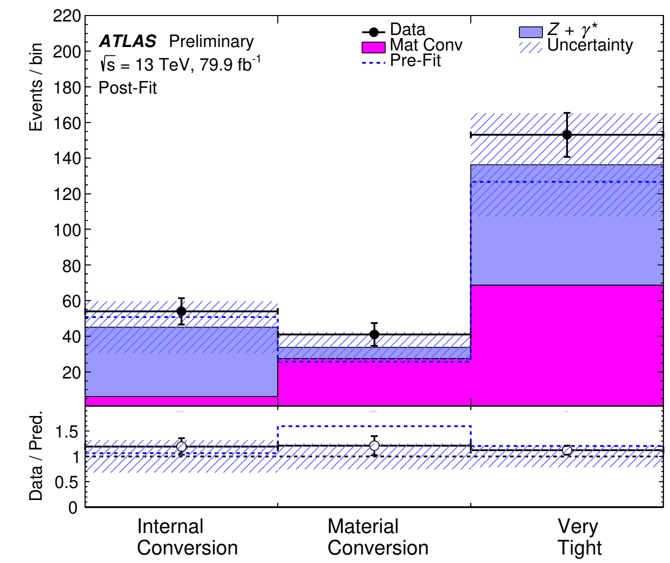
Measure in QCD di-jet events



Apply in the signal region



Plots from D. Dobur

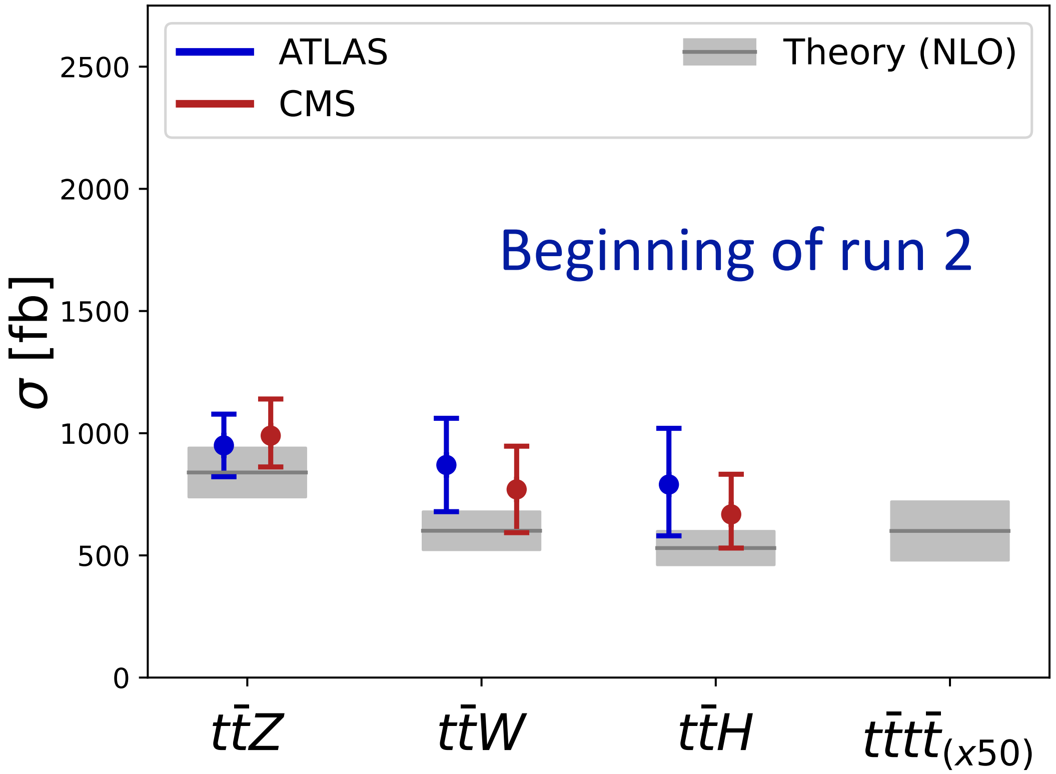


Parameter	$NF_{t\bar{t}W}$	NF_{CO}	NF_{γ^*}	NF_{HFe}	$NF_{HF\mu}$
Value	1.6 ± 0.3	1.6 ± 0.5	0.9 ± 0.4	0.8 ± 0.4	1.0 ± 0.4

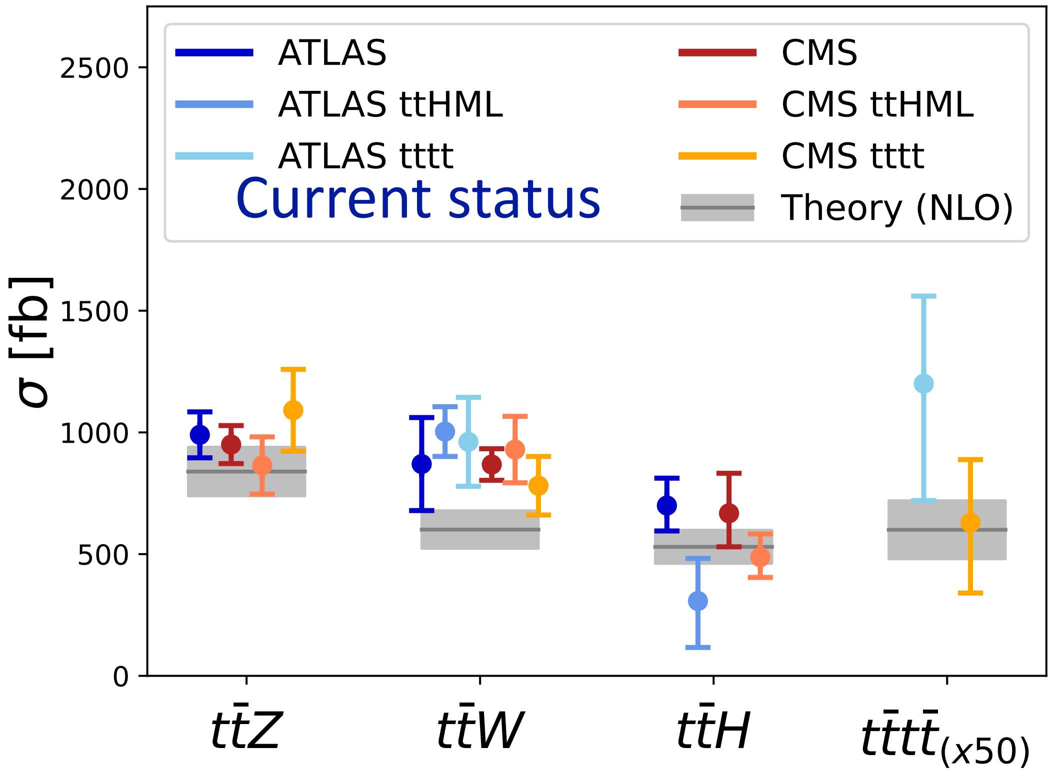
Given very different methods to estimate fakes by ATLAS and CMS it is unlikely that this is the source of the problem

Conclusions

- Majority of measurements in multilepton+b-jet final state use full run 2 data set
- Consistent picture between ATLAS and CMS
  - ttZ is consistent with the prediction within 1 sigma and does not show obvious mismodelling of kinematics
  - ttW normalisation is significantly above the prediction and there are hints of mismodelling on kinematics
  - differential measurement of ttW is critical to improve understanding of this background
    - uncertainties might be large: hard to define region pure in ttW with good statistics



Plots from Josh McFayden

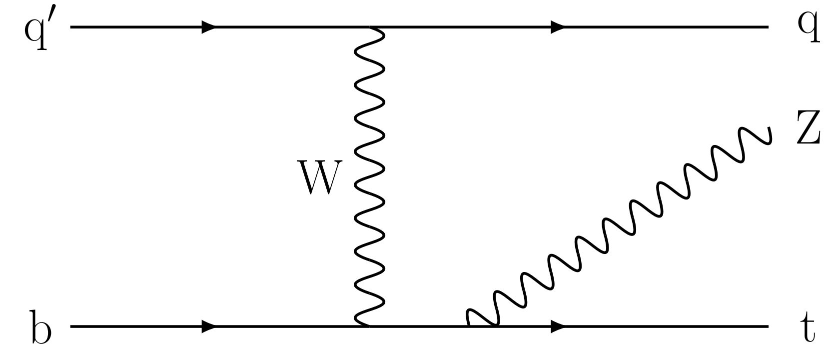


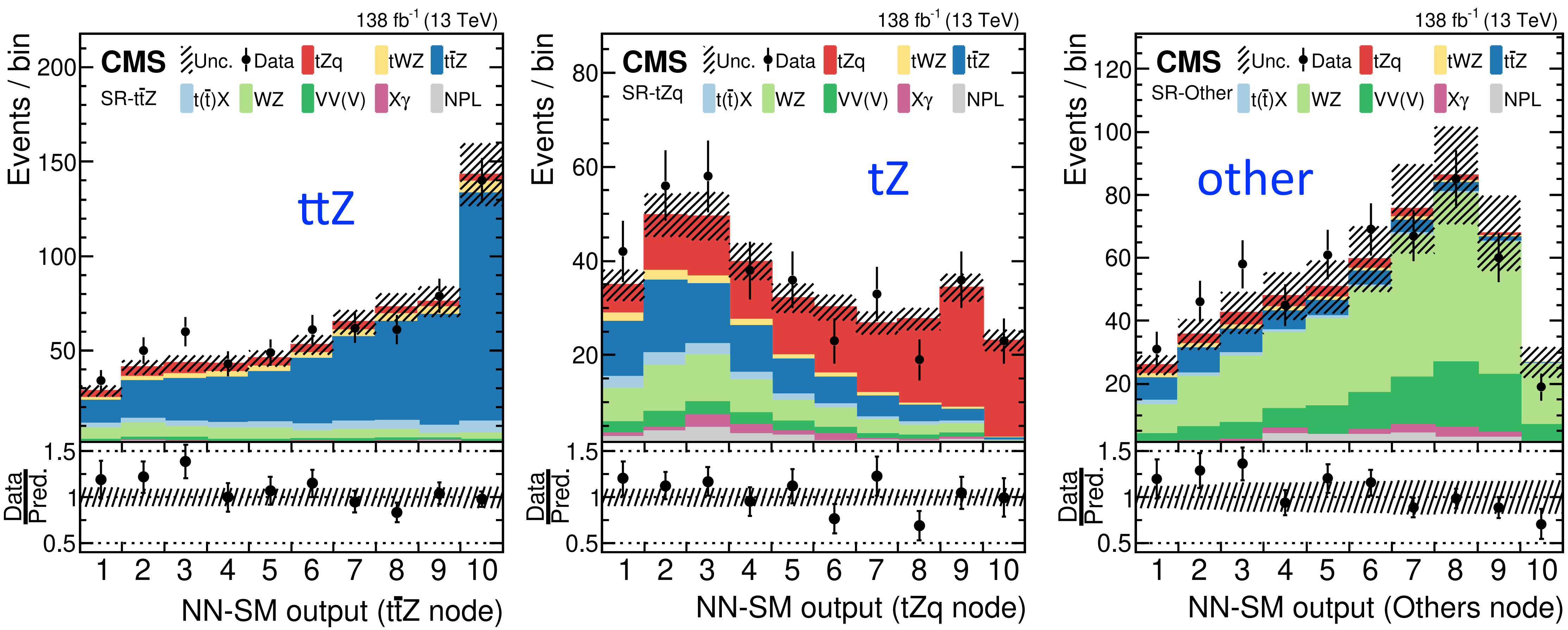
Backup

More ttZ

JHEP 12 (2021) 083

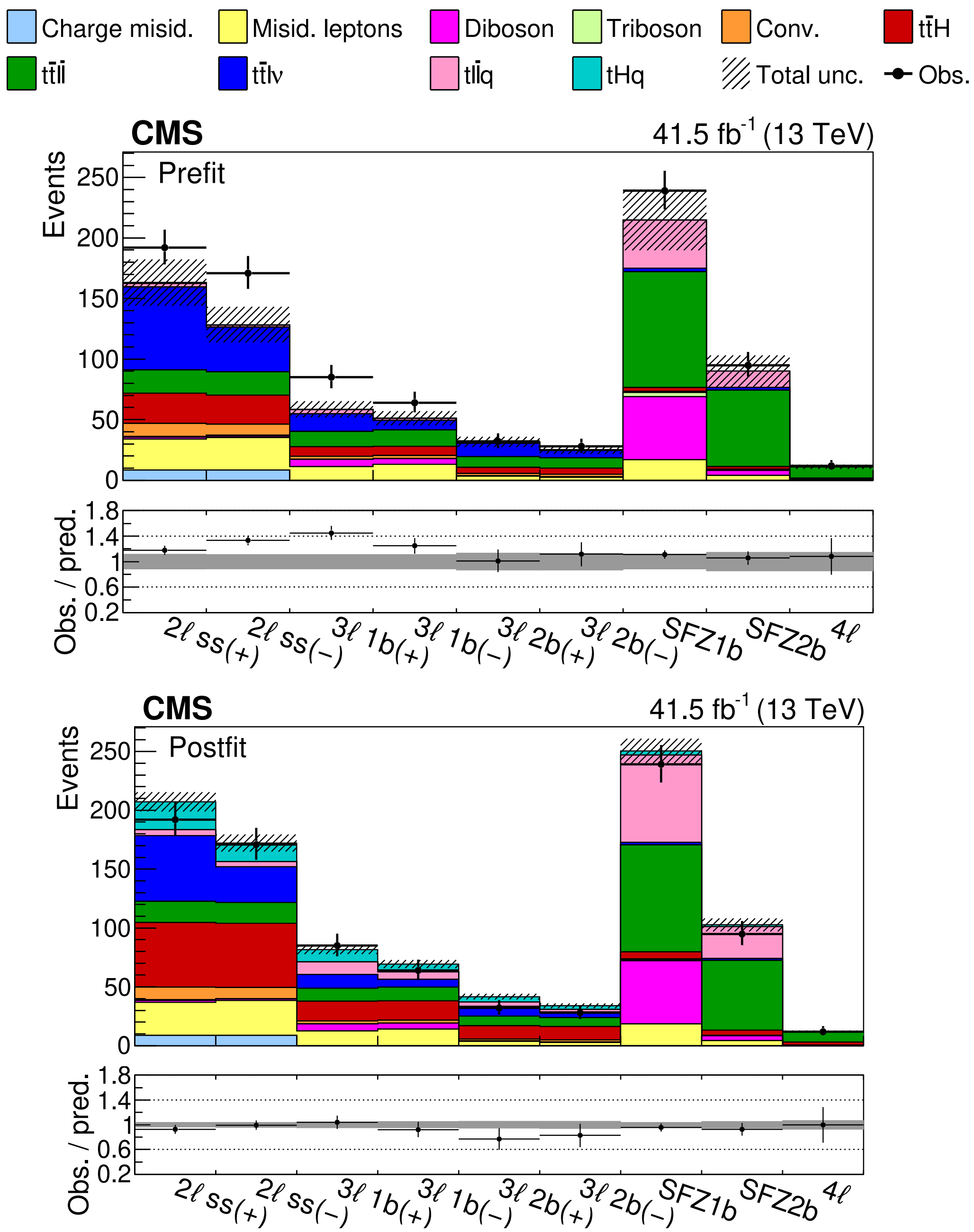
- EFT-focused analysis of ttZ and tZ events
- Full run 2 data set
- Same inclusive selection
- MVA to separate regions enriched in ttZ, tZ and other (mainly WZ)





Pre-fit plot don't seem to show any problem with ttZ normalisation

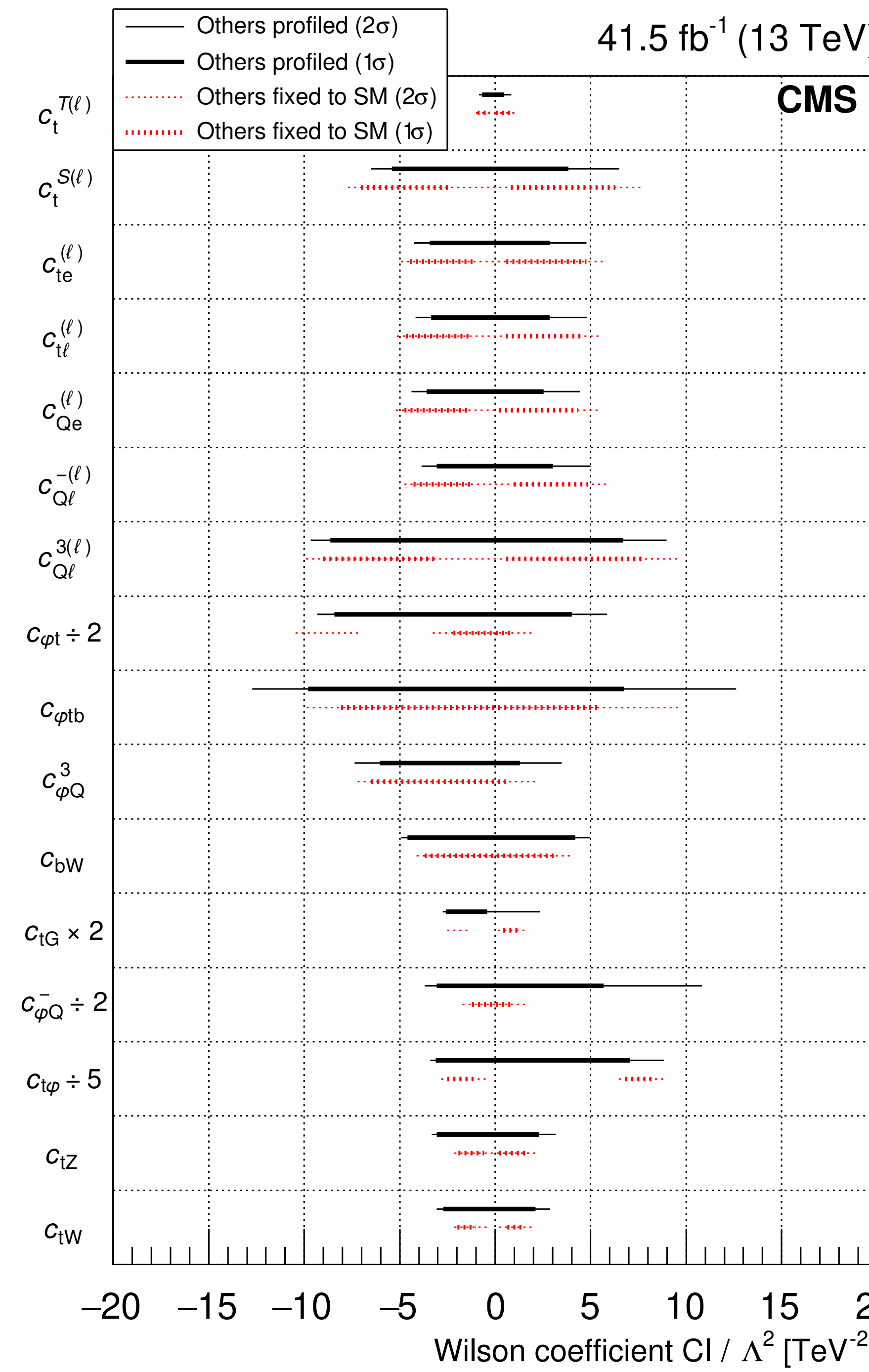
CMS search for new physics



prefit

postfit

JHEP 03 (2021) 095



- Consider 5 processes: ttH, tHq, ttℓℓ, tℓℓq, ttν
- SM constraints on all background normalisations
- Extraction of various Wilson coefficients