Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

INCLUSIVE $t\bar{t}$ CROSS SECTION IN THE ℓ+JETS CHANNEL AT 5.02 TEV WITH THE 2017 DATA SET

25/04/2024

Javier del Riego

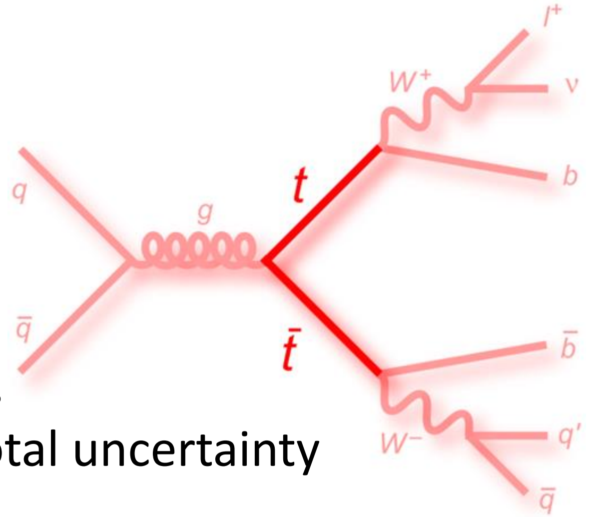
(on behalf of the $t\bar{t}$@5.02 TeV team)

Universidad de Oviedo - Universitè Catholique de Louvain - University of Kansas

LHC Top Working Group

Motivation

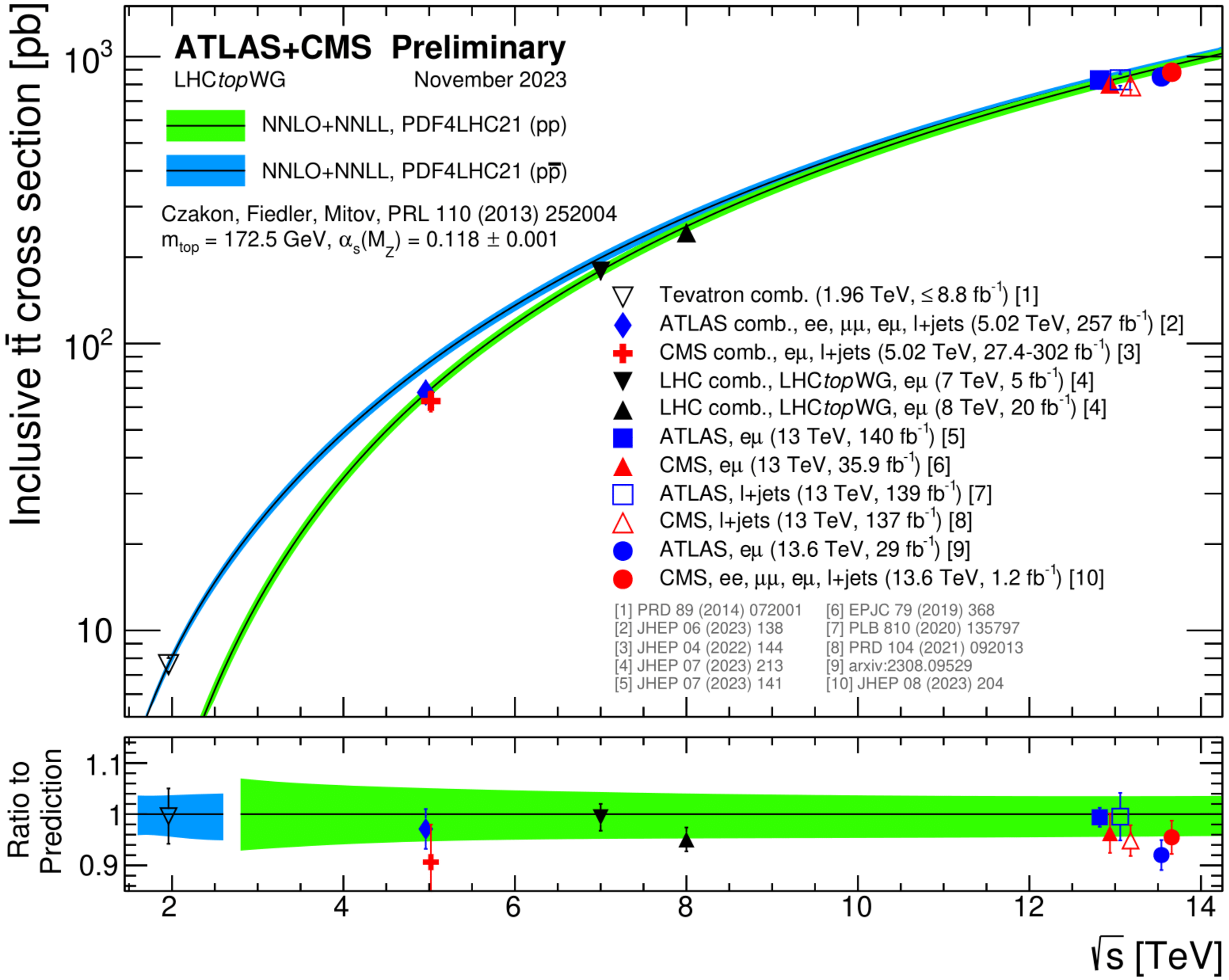
Goal: measure the $t\bar{t}$ **cross section at 5.02 TeV** in the semileptonic final state with the 2017 data 302 pb^{-1} .

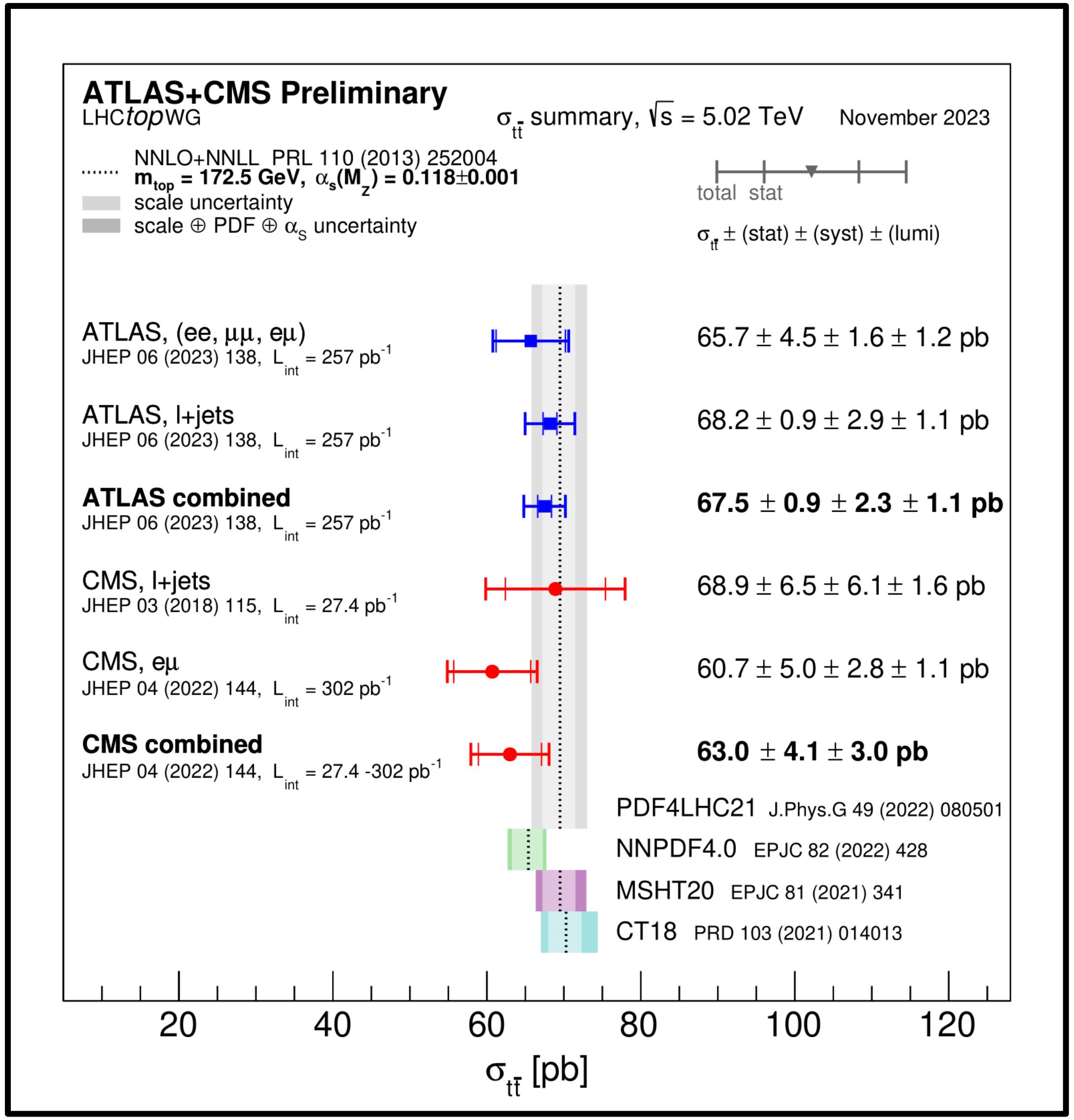


- **Previous measurements** overview:
  - JHEP 03 (2018) 115 (TOP-16-023): First measurement with 27 pb^{-1} (2015 data). Combination of **dilepton and single lepton** final states. Statistically dominated. Total uncertainty of 12%.
  - JHEP 04 (2022) 144 (TOP-20-004): Recent measurement with 302 pb^{-1} in the **eμ final state**. Uncertainty still dominated by statistics. **Combined with ℓ+jets from 2015** to improve the result. Total uncertainty of 8%.
  - JHEP 06 (2023) 138 (ATLAS): Combination **dilepton+single lepton** using the 2017 dataset. Uncertainty of about 4%.
    - Proposed at LHCTopWG to be combined with CMS final result.
- **We combine this new result with the dilepton analysis** (TOP-20-004) expecting to reduce the uncertainty, as the previous measurements are dominated by statistics.

Motivation

Goal: measure the $t\bar{t}$ **cross section at 5.02 TeV** in the semileptonic final state with the 2017 data 302 pb^{-1} .





SAMPLES/CORRECTIONS

Same as in TOP-20-004

➢ DATA:

- **2017** data, for an integrated luminosity of **302 pb^{-1}**
- Usage of **single-lepton** triggers
- Mean number of pp interactions per bunch crossing ≈ **2 -> low PU.**

➢ MC:

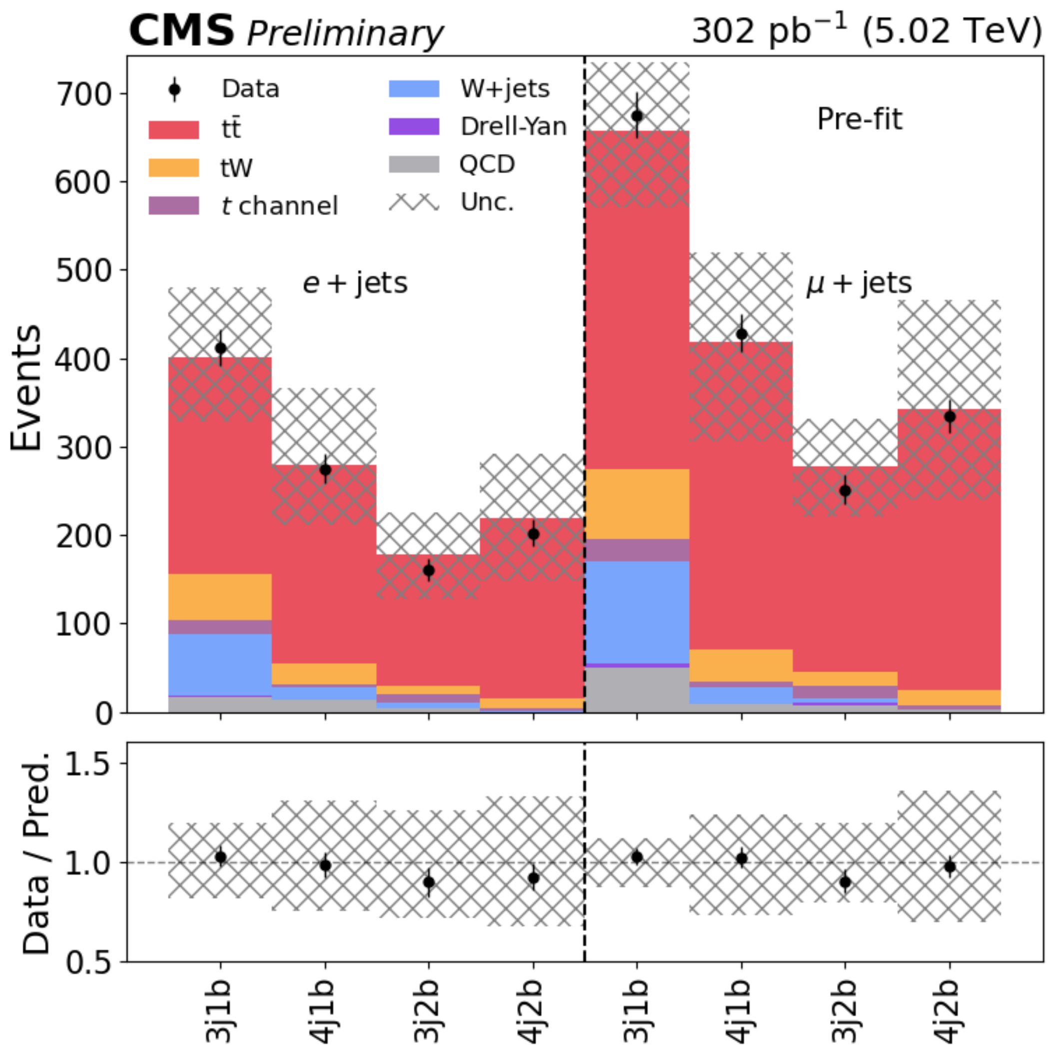
Process	Generator + Parton Shower	Cross section, σ_{norm} (pb)	Order of σ_{norm} approximation
$t\bar{t}$	POWHEG + PYTHIA 8	69.5	NNLO+NNLL [28, 29]
t channel	POWHEG + PYTHIA 8	30.3	Approximate NNLO [37]
tW	POWHEG + PYTHIA 8	6.54	Approximate NNLO [38]
W+jets	MADGRAPH5_aMC@NLO + PYTHIA 8	21159	NNLO[QCD]+NLO[EW] [39]
Drell-Yan	MADGRAPH5_aMC@NLO + PYTHIA 8	3647	NNLO[QCD]+NLO[EW] [39]

➢ Corrections:

- **Lepton** (e and μ), **Trigger, B-tag SFs + JECs.**

EVENT SELECTION

- **Exactly 1 lepton (electron or muon)** (p_T > 20 GeV, |η| < 2.4). Veto on sub-leading lepton of opposite flavour, p_T > 10 GeV.
- **At least 3 jets** (p_T > 25 GeV, |η| < 2.4).
- **MET > 30 GeV.**
- Events are further categorized into **8 categories** depending on the **number of jets and b-tagged jets, and the lepton flavour** (electron or muon). Among those:
  - All are signal-dominated
  - $\ell + 3j \geq 2b$ and $\ell + 4j \geq 2b$ are purest in signal (89% of total MC)
  - $\ell + 3j1b$ provide the greatest contribution from tW and W+jets backgrounds (12% and 18% of total MC)



Background estimation

- **3 main backgrounds:**
  - **Single top (tW + t-channel)**: main irreducible background. Estimated from MC.
  - **W+jets**: main background in 3j1b. Estimated from MC.
  - **QCD multijets**: mainly in some low b-tag categories (1b). Cut in **MET > 30GeV** to suppress it. Estimated from data in a region with non-isolated leptons and extrapolated to the signal region (see next slide). Following previous CMS analyses strategy.
- **Minor background:**
  - **Drell-Yan:** estimated from MC.

QCD ESTIMATION

1) Define a control region: “non-iso”

We invert the lepton MVA and the isolation requirements to obtain a region enriched in QCD events.

Estimation of QCD in the CR as: $N^{non\text{-}iso}(QCD) = N^{non\text{-}iso}(obs) - N^{non\text{-}iso}(MC)$

2) Calculate the extrapolation factor using low-MET events: MET < 20 GeV

Assuming that the reconstruction of a QCD lepton does not depend on MET, we compute the extrapolation factor from the control region to our signal region using low-MET events.

3) QCD estimation in the signal region:

$$N^{SR}(QCD) = \left(N^{non-iso}(obs) - N^{non-iso}(MC)\right) \times \frac{N^{SR}_{\text{low } \not{E}_T}(obs) - N^{SR}_{\text{low } \not{E}_T}(MC)}{N^{non-iso}_{\text{low } \not{E}_T}(obs) - N^{non-iso}_{\text{low } \not{E}_T}(MC)}$$

ANALYSIS STRATEGY

- Different observables were tested: m_T , m(j,j'), ΔR(j, j'), m(b,ℓ)…
- Finally, **median(ΔR(j, j'))** shape is used in the fits in every region expect the **3j1b category**, where an **MVA** is trained to further separate $t\bar{t}$ from W+jets.

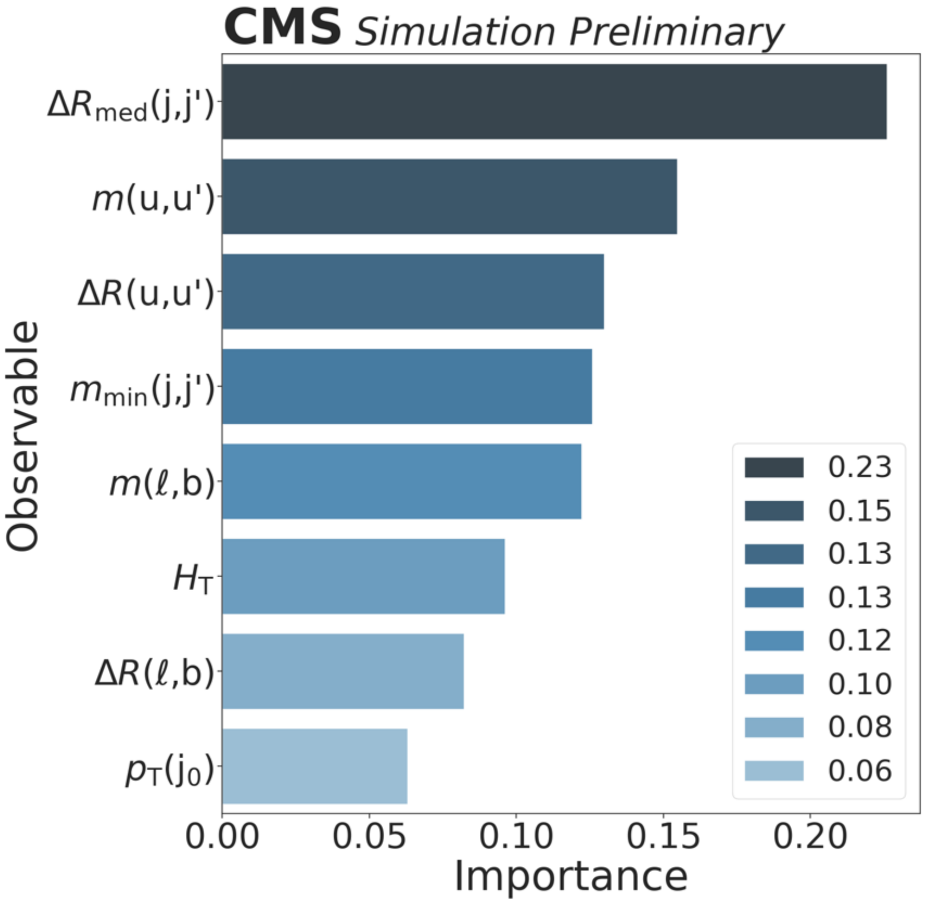
ANALYSIS STRATEGY

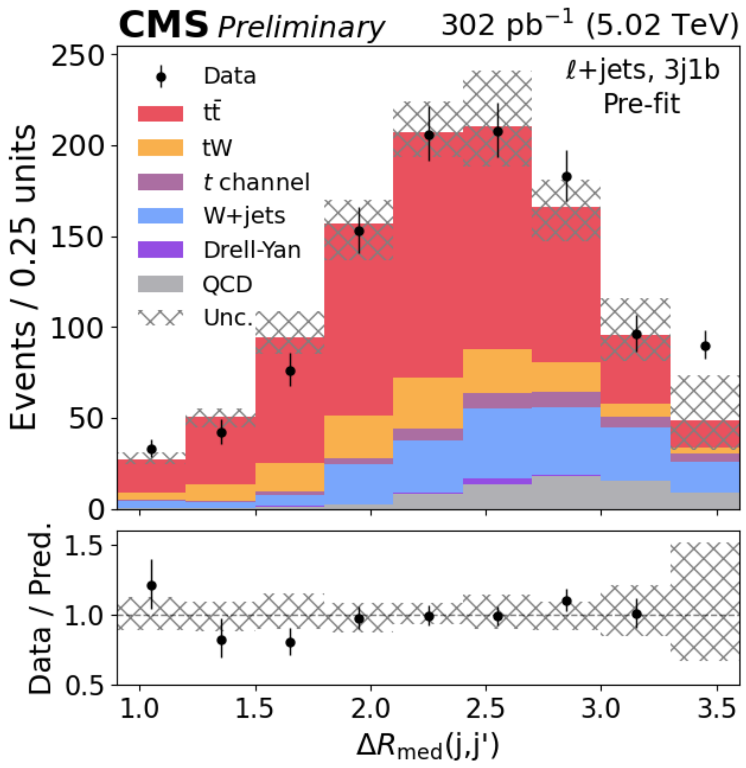
- Different observables were tested: m_T , m(j,j'), ΔR(j, j'), m(b,ℓ)…
- Finally, **median(ΔR(j, j'))** shape is used in the fits in every region expect the **3j1b category**, where an **MVA** is trained to further separate $t\bar{t}$ from W+jets.
- **MVA details:**
  - Model: random forest trained with Sklearn. 500 trees with max depth 6.
  - Signal: $t\bar{t}$ sample. Background: W+jets sample jet-binned.
  - Division of samples: 70% for train and 30% for test
  - 8 input variables: H_T, $j_0 p_t$, min m(j,j'), median ΔR(j, j'), m(b,ℓ), ΔR(b,ℓ), ΔR(u,u'), m(u,u').

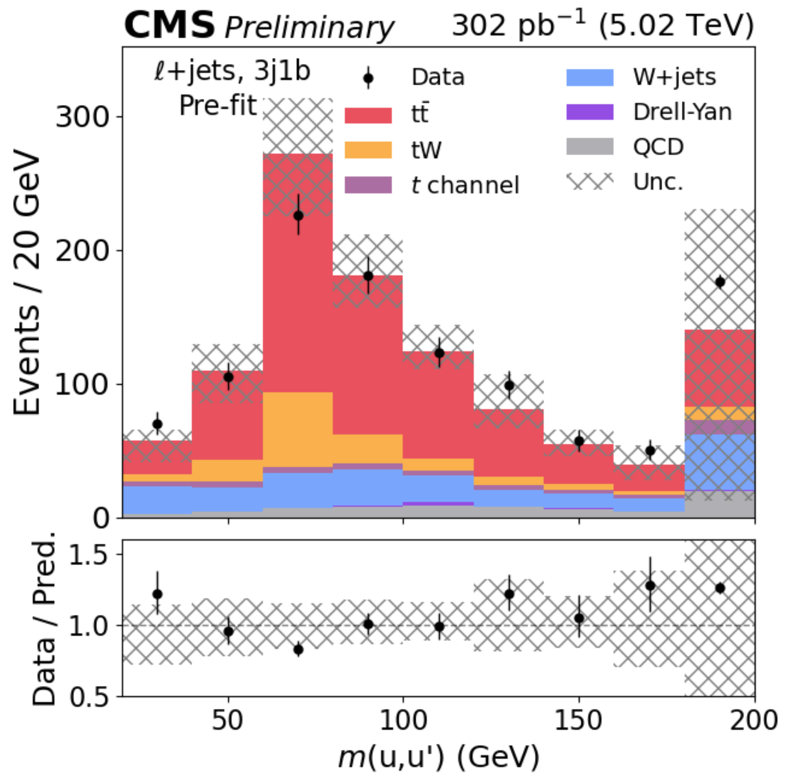
Variable	Definition
H_T	Scalar sum of the p_T of all jets in the event
$p_T(j_0)$	p_T of the leading jet
$m(u, u')$	Invariant mass of the two non-b-tagged jets
$\Delta R(u, u')$	ΔR between the two non-b-tagged jets
$m_{min}(j, j')$	Minimum invariant mass of all possible combinations of two jets
$\Delta R_{med}(j, j')$	Median ΔR between all possible combinations of two jets
$\Delta R(\ell, b)$	ΔR between the lepton and the b-tagged jet
$m(\ell, b)$	Invariant mass of the lepton and the b-tagged jet

ANALYSIS STRATEGY: MVA IN 3J1B

➢ The most important variables:

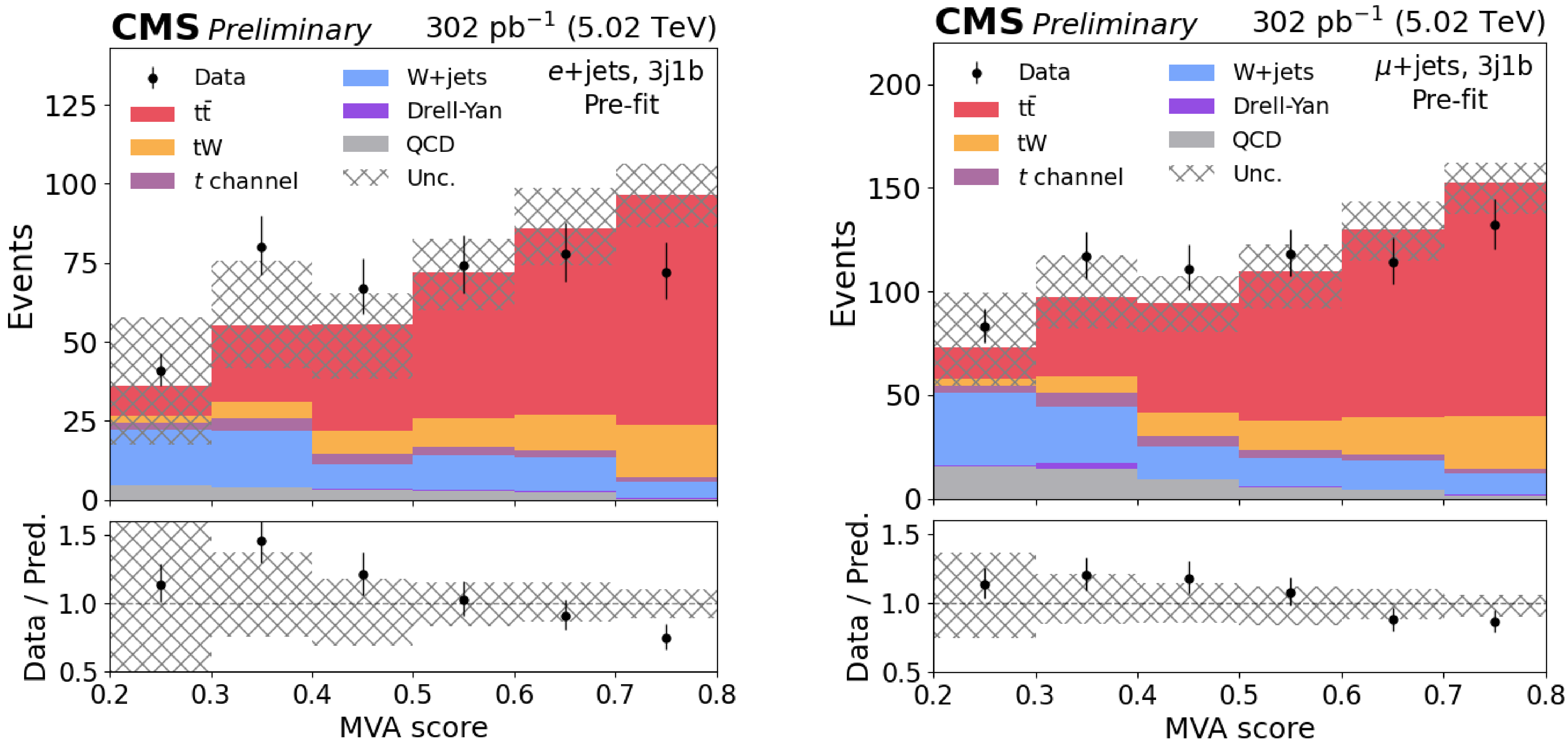






ANALYSIS STRATEGY: MVA IN 3J1B

➢ Obtaining the following discriminants:

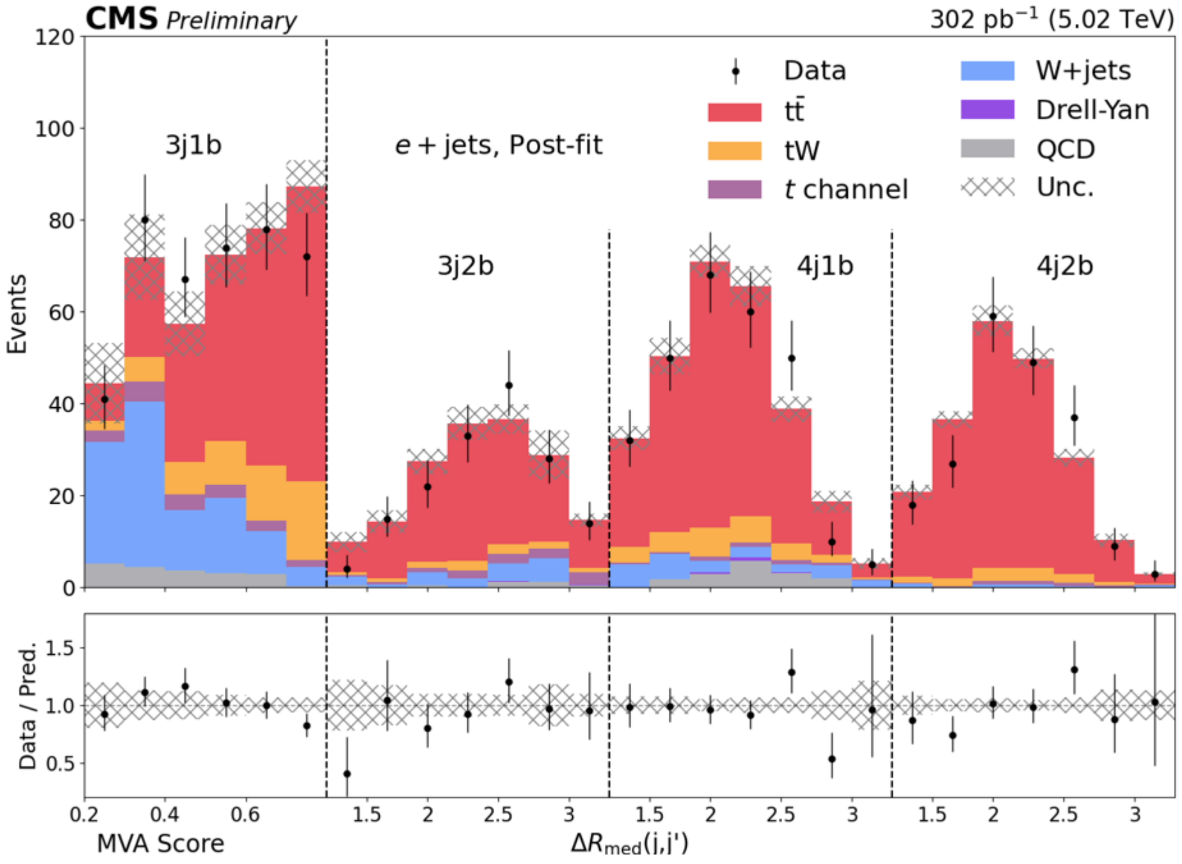


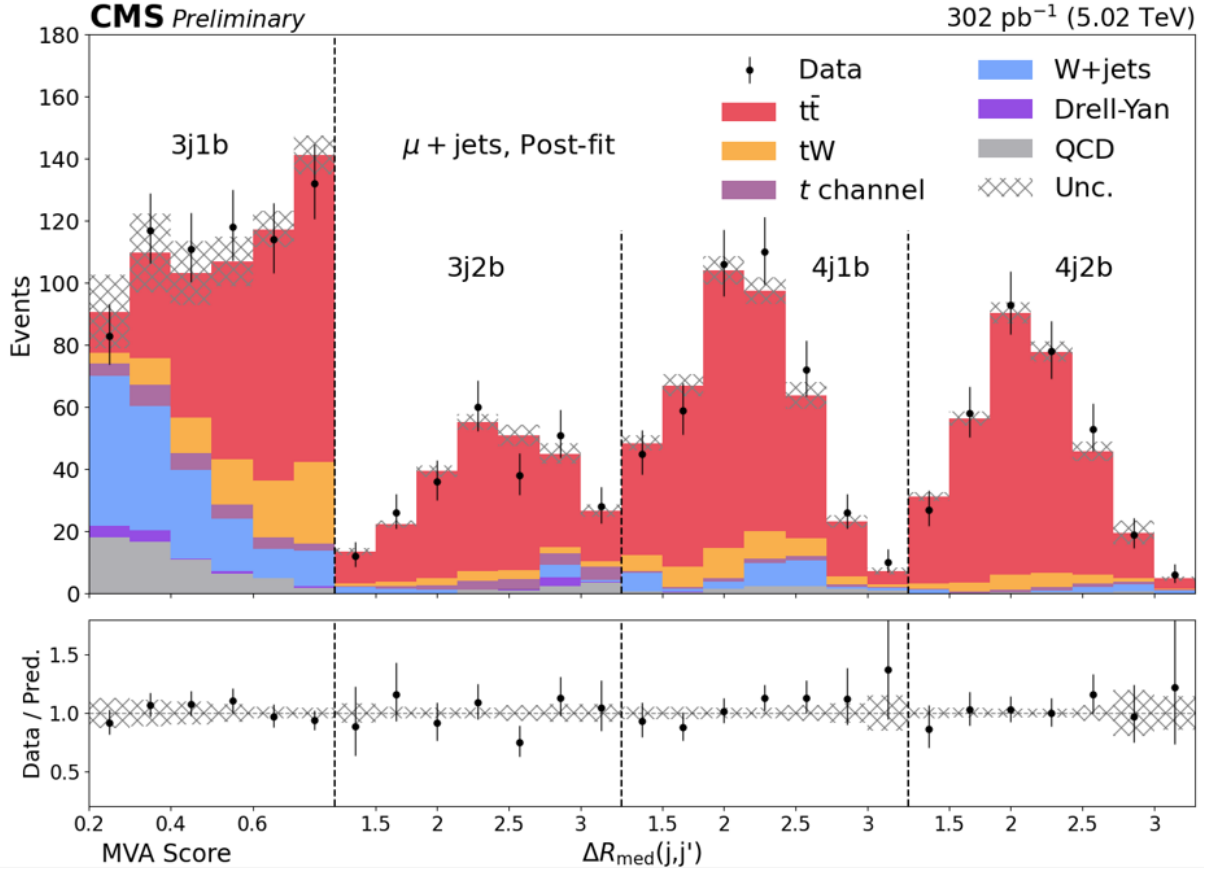
Cross section measurement

- Final distribution of 27 bins x 2 (e/μ) = 54 bins:

  median(ΔR(j, j')) + MVA Score (3j1b category)

- A maximum-likelihood fit is done simultaneously to the distributions.





Postfit distributions

UNCERTAINTIES

- **Experimental**
  - Lepton efficiencies
  - Trigger efficiencies
  - B-tagging: light (udsg) and heavy (bc).
  - L1 Prefiring
  - JEC and JER (13 sources)
  - Unclustered energy.
  - Stat. unc. from limited size of the MC samples

- ***$t\bar{t}$* modeling**
  - ISR/FSR: vary PS scales by 2, 0.5 factors
  - Underlying event: vary CP5 tuning → Normalization
  - ME/PS matching: vary h_{damp} of Powheg
  - ME scales: μ_F and μ_R scales by factors of 0.5 and 2
  - PDFs and α_S: 100+2 sources

Dedicated samples (Underlying event, ME/PS matching)

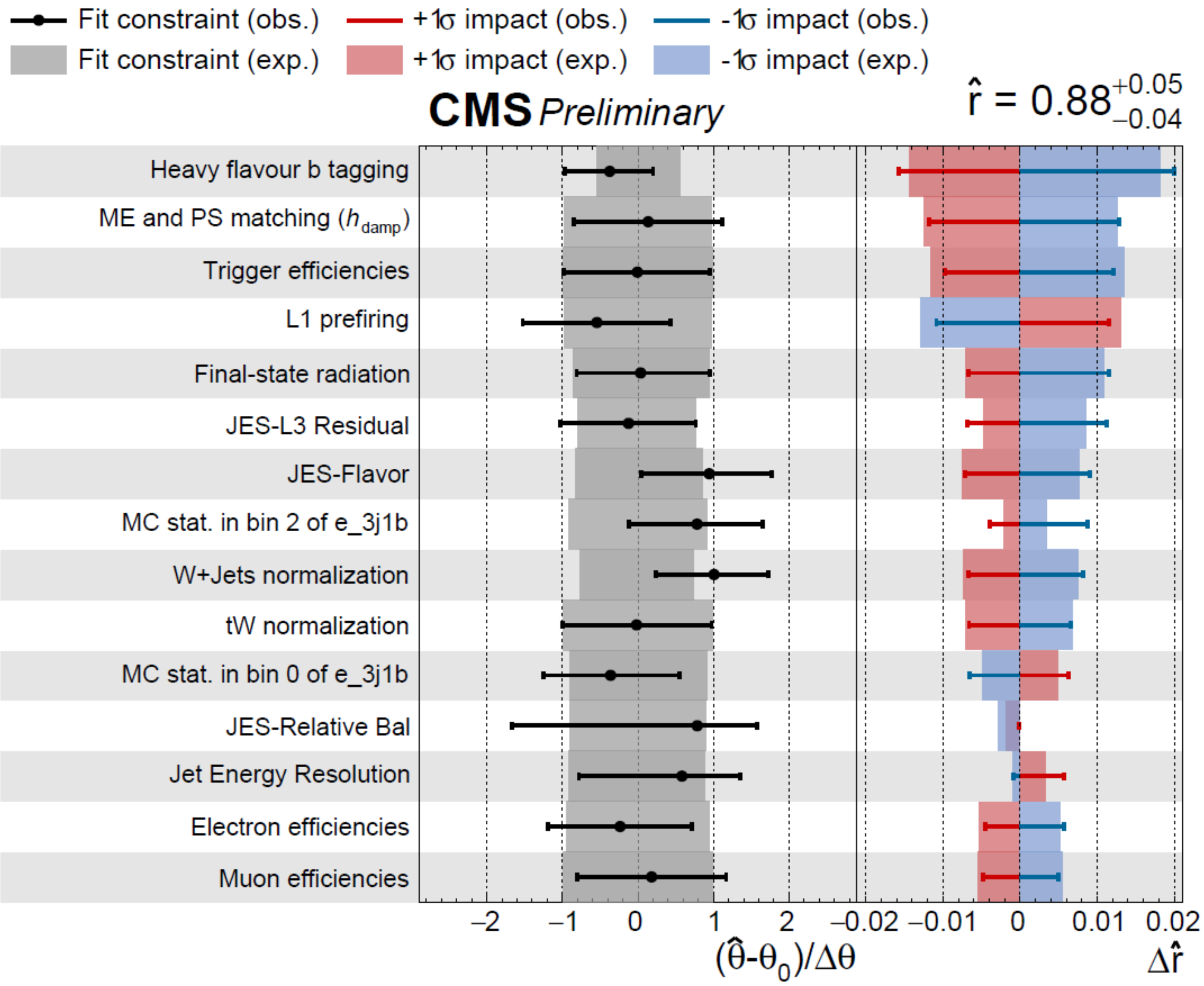
ΔR/MVA shape (Experimental and ISR/FSR)

Nj/Nb shape (ME/PS matching, ME scales, PDFs and α_S)

- **Normalization**
  - Background normalization -> QCD: stat+30% norm, DY: 30%, W+Jets: 20%, tW: 5.6%, t channel: 10%
  - Luminosity: 1.9%

OBSERVED RESULTS IN ℓ + JETS

- We observe a **total uncertainty of 5.5% and *r* = 0.88.** (Luminosity added afterwards)



- This results in a measured cross section of:

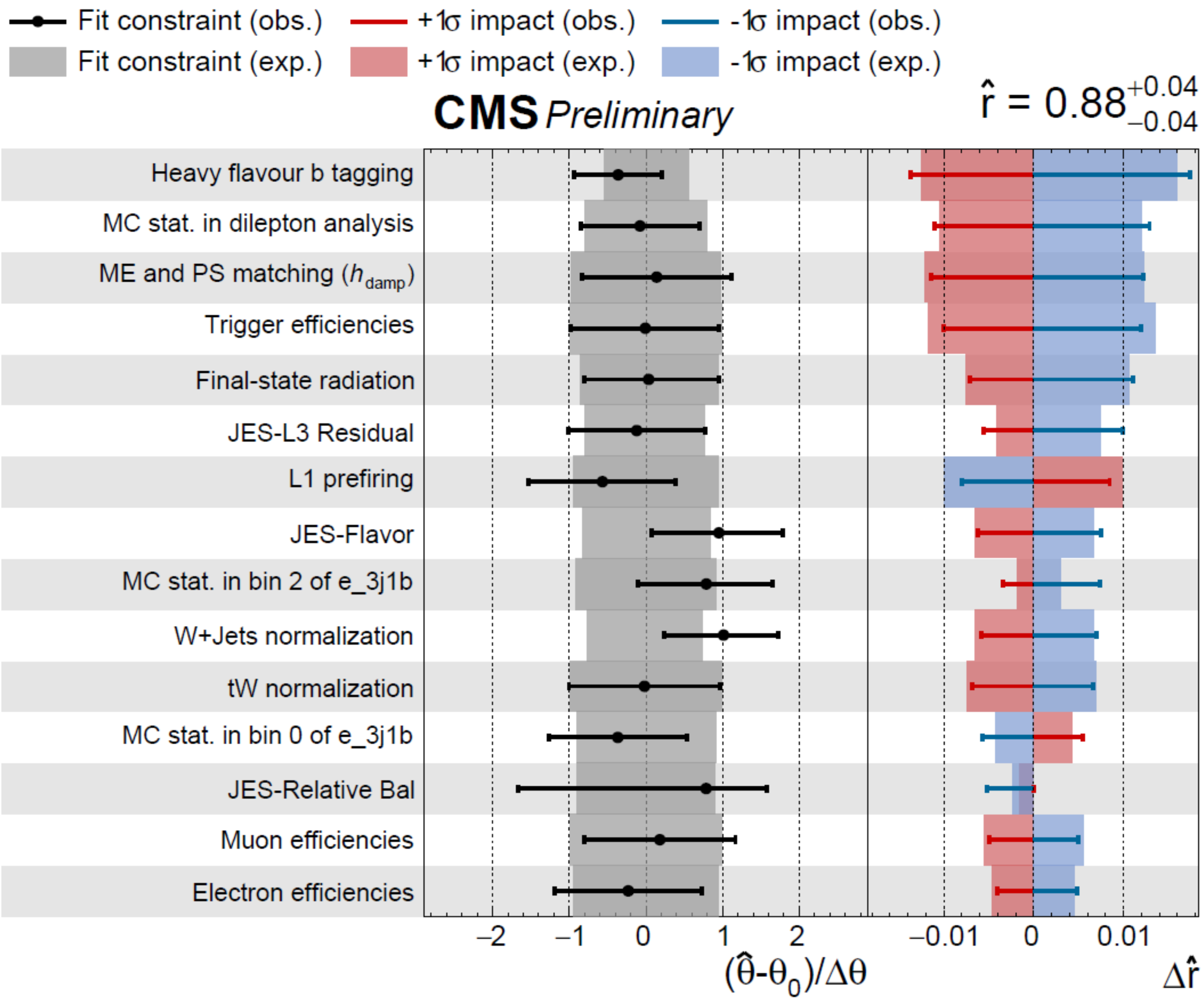
$$\sigma_{t\bar{t}} = 61.4 \pm 1.6\ (stat)^{+2.7}_{-2.6}(syst) \pm 1.2\ (lumi)\ pb$$

CORRELATION SCHEME FOR FINAL COMBINATION WITH DILEPTON RESULT

Source	Correlation with 2ℓ	
Electron efficiency	100%	
Muon efficiency	100%	
Trigger efficiency	100%	
b-tagging	0%	not included in dilep
JES	0%	in dilep 1 source, in ℓ+jets splitted
JER	100%	
Unclustered energy	0%	(not included in dilep)
L1 prefiring	100%	
Final-state radiation	100%	
Initial-state radiation	100%	
μ_R, μ_F scales	100%	
PDF$\oplus\alpha_S(m_Z)$	0%	in dilep 1 source, in ℓ+jets splitted
h_{damp}	100%	
Underlying event tune	100%	
tW	100%	
QCD	0%	not included in dilep
Drell–Yan	100%	
W+jets	100%	
t channel	0%	not included in dilep
Integrated luminosity	100%	

OBSERVED RESULTS COMBINED WITH DILEPTON ANALYSIS

- Combining with previous CMS dilepton result, **uncertainty from 5.5% to 5.1%** and r = 0.88.

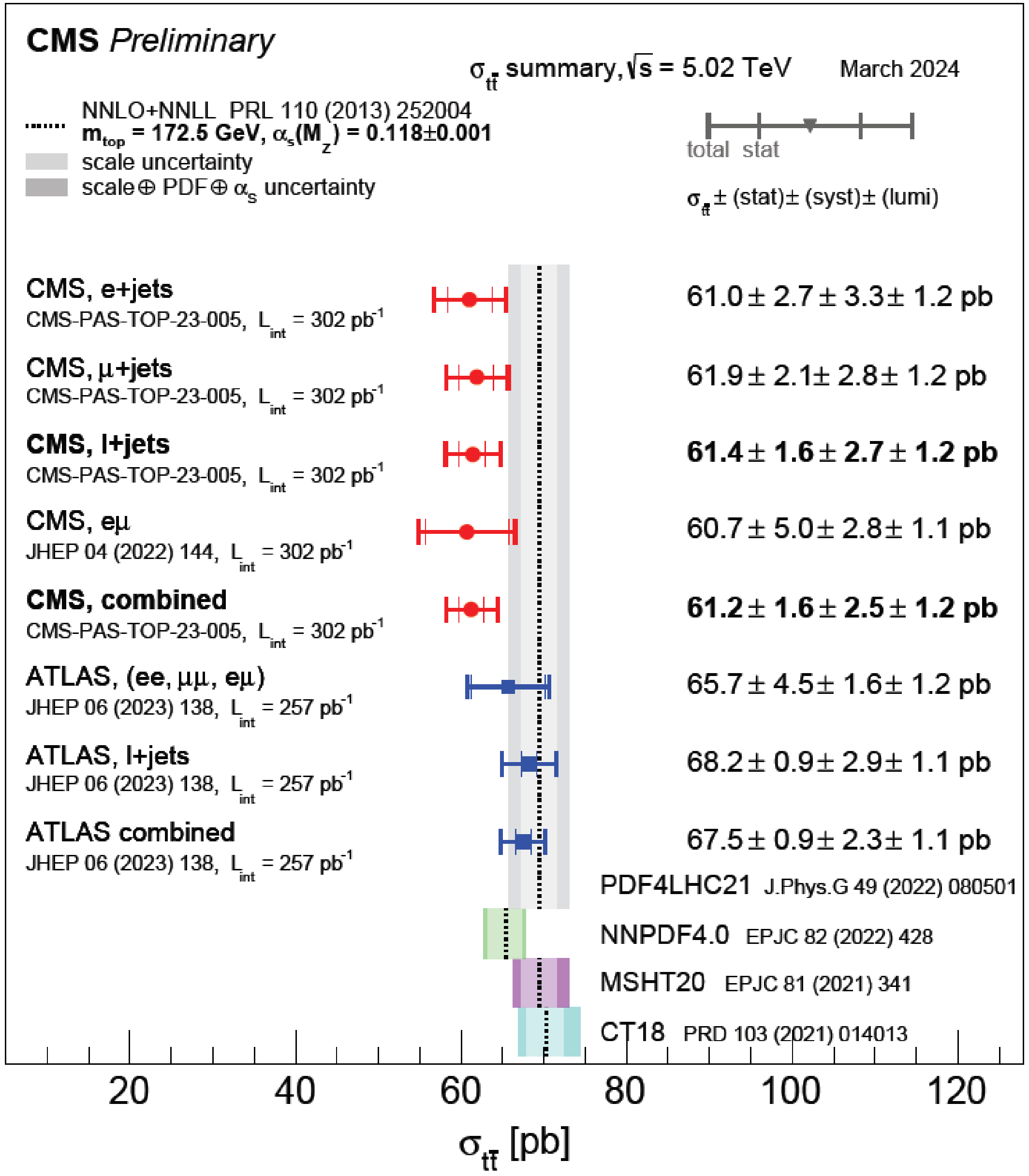


- This results in a measured cross section of:

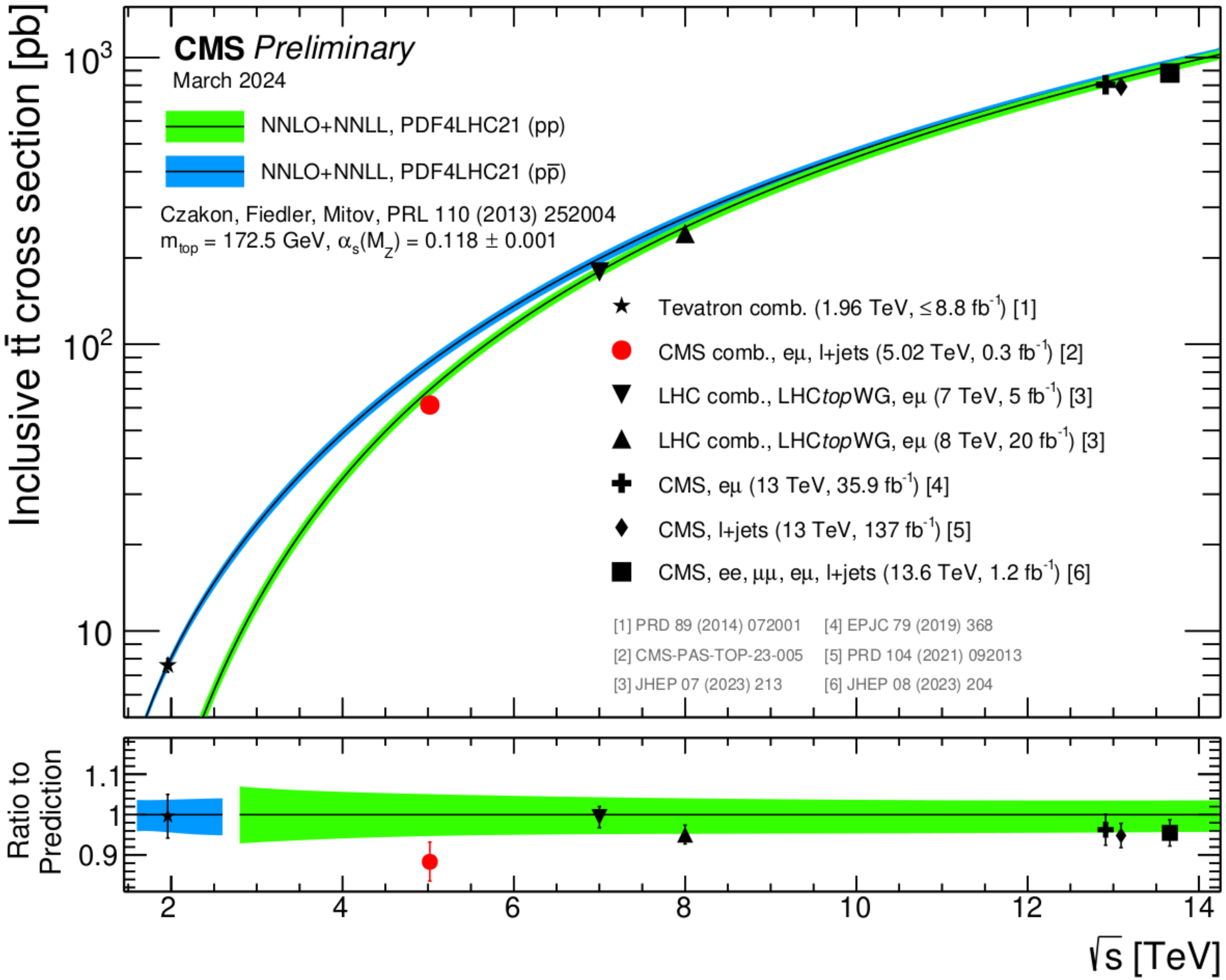
$$\sigma_{t\bar{t}} = 61.2^{+1.6}_{-1.5}(stat)^{+2.6}_{-2.3}(syst) \pm 1.2\ (lumi)\ pb$$

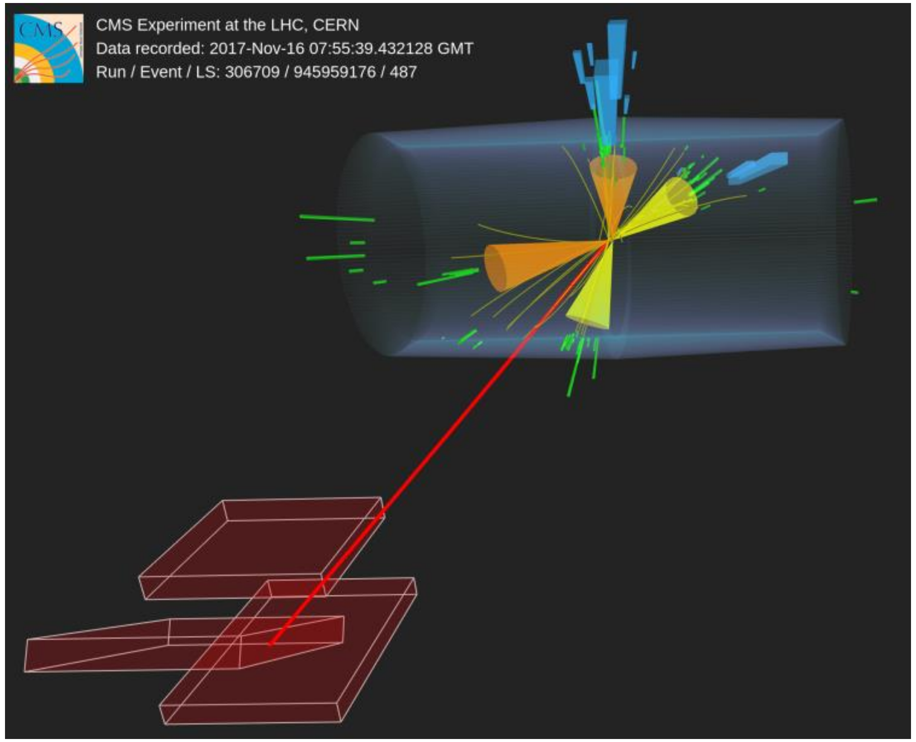
SUMMARY

- An analysis that means great improvement with respect to previous results is observed:
  - previous ℓ+jets CMS result (TOP-16-023): **13% → 5.5%.**
  - previous dilepton & ℓ+jets CMS result (TOP-20-004): **8.4% → 5.1%.**
- Most precise CMS result for the $t\bar{t}$ inclusive cross section at 5.02 TeV! This would help to further improve the uncertainty combining with the ATLAS measurement [JHEP 06 (2023) 138], since we have improved the previous CMS systematic precision to an extent compatible with ATLAS.
- Recently become public: PAS-TOP-23-005



Thanks for the attention!





LHCtopWG

Back up

Summary of previous results

		Stat.	Syst.	Lumi	Total
Dilep 2015	CMS	24.7%	5.2%	2.6%	25%
Semilep 2015	CMS	9.4%	8.8%	2.3%	13.1%
Comb. dilep+semilep 2015	CMS	8.8%	8%	2.3%	12.1%
Semilep 2017	CMS	2.6%	4.4%	1.9%	5.5%
	ATLAS	1.3%	4.3%	1.6%	4.5%
Dilep 2017	CMS	8.2%	4.6%	1.9%	9.6%
	ATLAS	6.8%	2.5%	1.8%	7.5%
Comb. dilep 2017 + semilep 2015	CMS	6.5%	4.7%		8.4%
Comb. dilep + semilep 2017	CMS	2.4%	4.1%	1.9%	5.1%
	ATLAS	1.3%	3.4%	1.6%	3.9%

Comparison with ATLAS result

- **Uncertainties in l+jets analysis:**
  ATLAS: 1.3 (stat) +- 4.3 (syst) +- 1.6 (lumi) %
  CMS: 2.6 (stat) +- 4.4 (syst) +- 1.9 (lumi) %

- We **believe** that ATLAS is not considering the background stat. uncertainty in the computation of the total stat. uncertainty.

ATLAS

	$\ell+2j \geq 1b$	$\ell+3j\ 1b$	$\ell+3j\ 2b$	$\ell+\geq 4j\ 1b$	$\ell+4j\ 2b$	$\ell+\geq 5j\ 2b$
$t\bar{t}$	194 ± 27	310 ± 33	199 ± 24	690 ± 60	318 ± 32	380 ± 60
Single top	195 ± 22	98 ± 12	38 ± 5	67 ± 9	22 ± 4	15.9 ± 2.7
W+jets	1700 ± 400	690 ± 210	58 ± 23	350 ± 120	30 ± 14	19 ± 10
Other bkg.	110 ± 40	55 ± 23	7.2 ± 3.0	29 ± 12	3.5 ± 1.5	3.7 ± 1.7
Misidentified leptons	250 ± 130	110 ± 60	10 ± 5	60 ± 30	6 ± 3	8 ± 5
Total	2500 ± 400	1260 ± 210	312 ± 34	1200 ± 160	380 ± 40	430 ± 70
Data	2411	1214	293	1135	375	444

Data: 5872 **Bkg: 3991**

$$\%\ stat. = \frac{\sqrt{data}}{data - bkg}$$

Bkg=0 → **1.3%** → **4.1%**

CMS

Yields	$\ell+3j1b$	$\ell+\geq 4j1b$	$\ell+3j\geq 2b$	$\ell+\geq 4j\geq 2b$
tt	627.6	573.4	381.8	522.3
tW	131.4	59.5	24.1	28.7
tchan	40.3	9.2	25.1	7.9
DY	5.3	0.2	3.0	0.0
WJets	185.9	33.6	11.5	1.0
QCD	67.5	22.2	10.6	2.0
total	1058.0	698.2	456.2	562.0
data	1087.0	703.0	411.0	536.0

Data: 2737 **Bkg: 669.3**

$$\%\ stat. = \frac{\sqrt{data}}{data - bkg}$$

Bkg=0 → **1.9%** → **2.6%**

OBJECT SELECTION AND TRIGGERS

- **Leptons**: Lepton MVA ID from ttH analysis [arXiv:2310.03844]. **Same as in TOP-20-004**
  $p_T > 20$ GeV, $|\eta| < 2.4$.

- **Jets:** AK4PFchs jets, Tight ID cleaned with selected leptons. **Same as in TOP-20-004**
  $p_T > 25$ GeV, $|\eta| < 2.4$.

- **B-tagged jets:** Medium WP of the DeepCSV algorithm.

- **MET**: using type-1 corrected MET.

- **Triggers**: unprescaled single lepton triggers. **Same as in TOP-20-004**

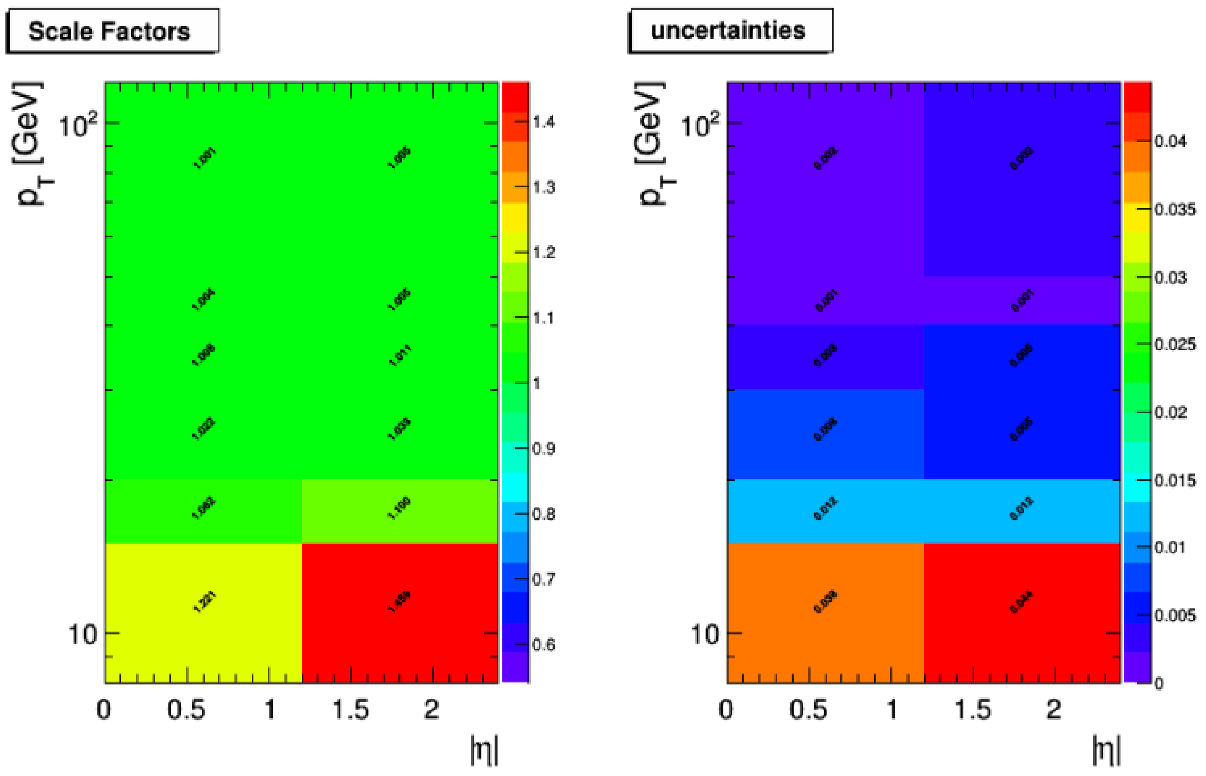
CORRECTIONS

Same as in TOP-20-004

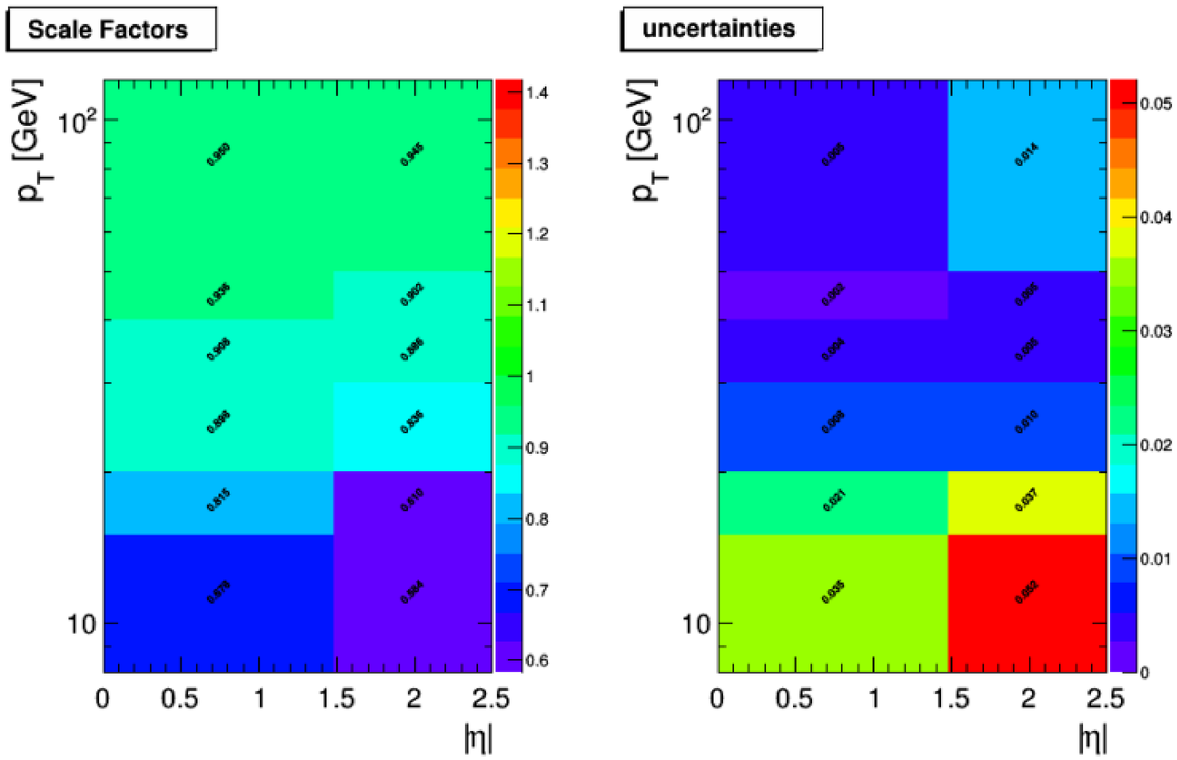
➢ Lepton SFs

- Estimated with the Tag & Probe method, which has been used many times previously.
- Used in other analyses at 5 TeV (TOP-20-004, SMP-20-012) and documented in AN-2020/093.

Loose-tight muon SFs



Loose-tight electron SFs



CORRECTIONS

➢ Trigger SFs

- Estimated using the **cross-trigger method**.
- Trigger efficiencies ≈ 95% for muons and ≈ 85% for electrons.
- Applied in 2 η bins.

$$Eff_{muon} = \frac{Events\ passing\ elec\ \&\ muon\ trigger}{Events\ passing\ elec\ trigger}$$

$$Eff_{elec} = \frac{Events\ passing\ elec\ \&\ muon\ trigger}{Events\ passing\ muon\ trigger}$$

⇩

$$SF = \frac{Eff_{data}}{Eff_{MC}}$$

CORRECTIONS

Same as in TOP-20-004

- **Jet energy corrections (JECs)**
  - Spring18_ppRef5TeV_V4 jet energy corrections are used.
  - Dedicated jet energy scale (JES) and jet energy resolution (JER) for this Run derived by HIN, in this twiki.
  - These corrections are also applied in MET.
- **B-tagging SFs**
  - 94X efficiencies derived for 2017 13 TeV dataset (same detector conditions).
  - MC efficiencies derived with low PU (5.02 TeV) $t\bar{t}$ sample.
  - ‘mujet’ method is used.
  - Divided into light (udgs) and heavy (bc) flavour.

Experimental uncertainties

- **Lepton efficiencies**: varying dedicated SFs by their uncertainties.
- **Trigger efficiencies**: propagating the efficiency uncertainty to the SFs.
- **B-tagging**: estimated by varying the b-tagging scale factors by the uncertainties. Splitted in light (udsg) and heavy (bc) flavours.
- **L1 Prefiring:** dedicated for 2017 Run G.
- **JEC and JER uncertainties:** the available 13 sources are taken into account separately (’MC’, ’Abs-Stat’, ’AbsScale’, ’AbsMPF’, ’Frag’, ’ECAL’, ’HCAL’, ’Flavor’, ’RelStat’, ’RelPt’, ’RelBal’, ’RelJER’, ’L3Res’).
- **Unclustered energy**: uncertainty in MET from the contribution of unclustered energy evaluated based on the momentum resolution of the different PF candidates.

These uncertainties are applied to all signal and background predictions.

MODELING $t\bar{t}$ UNCERTAINTIES

- **Initial and final state radiation (ISR/FSR)**: varying the parton shower scales by factors of 2 and 0.5 (weights).

- **ME/PS matching (h_{damp})**: varying the h_{damp} parameter within the uncertainties and propagating the result to the final yields (dedicated samples). Normalization in jet multiplicity bins.

- **Underlying event:** varying CP5 tune with uncertainties up and down (dedicated samples). Single uncertainty for all bins.

- **ME scales**: changing μ_F and μ_R scales by factors of 0.5 and 2 (weights), excluding variations of the scales in opposite directions. The envelope of all the variations is taken as uncertainty.

- **PDFs and α_S**: reweighting the $t\bar{t}$ sample according to 100 NNPDF3.1 NNLO PDF error sets and two extra weights for α_s (weights). All the sources are taken into account separately.

NORMALIZATION UNCERTAINTIES

- **Background normalization:**
  - **QCD:** stat. from the method + 30% normalization.
  - **DY**: 30%.
  - **W + jets**: 20%.
  - **tW**: 5.6%.
  - **t-channel**: 10%.

- **Luminosity**: 1.9% [ref]. Freezed on the fit and added quadratically afterwards.

These uncertainties have only an effect in the normalization.