IFCA Instituto de Física de Cantabria

EXCELENCIA MARÍA DE MAEZTU

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UC UNIVERSIDAD DE CANTABRIA

W^+W^- boson pair production at 13 TeV using CMS data

DIS2021: XXVIII International Workshop on Deep-Inelastic Scattering and Related Subjects

Pedro J. Fernández Manteca

IFCA (University of Cantabria, CSIC)

12-16 of April, 2021

CMS

DIS 2021

Introduction

- WW cross section measurements at 13 TeV using the CMS full 2016 dataset
- Two different approaches:
    - **Sequential** analysis: **SEQ**
    - **Random Forest** analysis (multivariate technique): **RF**
- **Goals:**
    - Measure the total WW cross section in the dileptonic channel (**SEQ** & **RF**)
    - Measure fiducial & differential cross sections in $m_{\ell\ell}$, $p_{T\ell 1}$, $p_{T\ell 2}$, $\Delta\phi_{\ell\ell}$ (**SEQ**) + njets (**RF**)
    - Set limits on Wilson coefficients (**SEQ**)

Phys. Rev. D 102, 092001 (2020)

High Energy Physics – Experiment

[Submitted on 31 Aug 2020]

W^+W^- boson pair production in proton–proton collisions at $\sqrt{s} = 13$ TeV

CMS Collaboration

A measurement of the W^+W^- boson pair production cross section in proton–proton collisions at $\sqrt{s} = 13$ TeV is presented. The data used in this study are collected with the CMS detector at the CERN LHC and correspond to an integrated luminosity of 35.9 fb^{-1}. The W^+W^- candidate events are selected by requiring two oppositely charged leptons (electrons or muons). Two methods for reducing background contributions are employed. In the first one, a sequence of requirements on kinematic quantities is applied allowing a measurement of the total production cross section: 117.6 ± 6.8 pb, which agrees well with the theoretical prediction. Fiducial cross sections are also reported for events with zero or one jet, and the change in the zero-jet fiducial cross section with the jet transverse momentum threshold is measured. Normalized differential cross sections are reported within the fiducial region. A second method for suppressing background contributions employs two random forest classifiers. The analysis based on this method includes a measurement of the total production cross section and also a measurement of the normalized jet multiplicity distribution in W^+W^- events. Finally, a dilepton invariant mass distribution is used to probe for physics beyond the standard model in the context of an effective field theory, and constraints on the presence of dimension-6 operators are derived.

Comments: Submitted to Phys. Rev. D. All figures and tables can be found at this http URL (CMS Public Pages)
Subjects: **High Energy Physics – Experiment (hep-ex)**

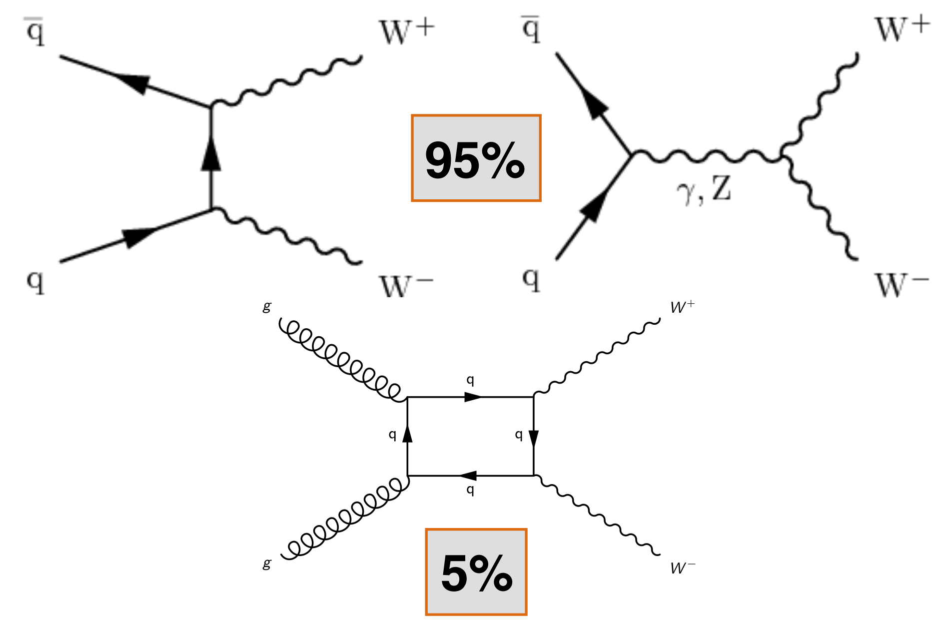
Last public result: CMS-PAS-SMP-16-006 (2.3 fb^{-1})
WW total Xsec = 115.3 ± 11.0 pb

This result: (35.9 fb^{-1})
WW total Xsec = 117.6 ± 6.8 pb

Sequential analysis

- The sequential selection relies mainly on a set of **discrete requirements** on kinematic variables and on a multivariate analysis tool to suppress Drell-Yan background in same-flavour channel

Target signature: two opposite charged isolated leptons, and large transverse missing energy from the neutrinos



Signal Region definition:

Quantity	Sequential Cut: Different-flavor	Sequential Cut: Same-flavor
Number of leptons	Strictly 2	
Lepton charges	Opposite	
$p_T^{\ell\,\max}$	>25	
$p_T^{\ell\,\min}$	>20	
$m_{\ell\ell}$	>20	>40
Additional leptons	0	
$\lvert m_{\ell\ell} - m_Z \rvert$	—	>15
$p_T^{\ell\ell}$	>30	>30
p_T^{miss}	>20	>55
* $p_T^{miss,proj}$, $p_T^{miss,track\ proj}$	>20	>20
Number of jets	≤1	
Number of b-tagged jets	0	
DYMVA score	—	>0.9

for DY suppression ($m_{\ell\ell}$ >40, $\lvert m_{\ell\ell} - m_Z \rvert$ >15, $p_T^{\ell\ell}$ >30, p_T^{miss} >55, DYMVA score >0.9)

suppress ttbar (Number of jets ≤1, Number of b-tagged jets 0)

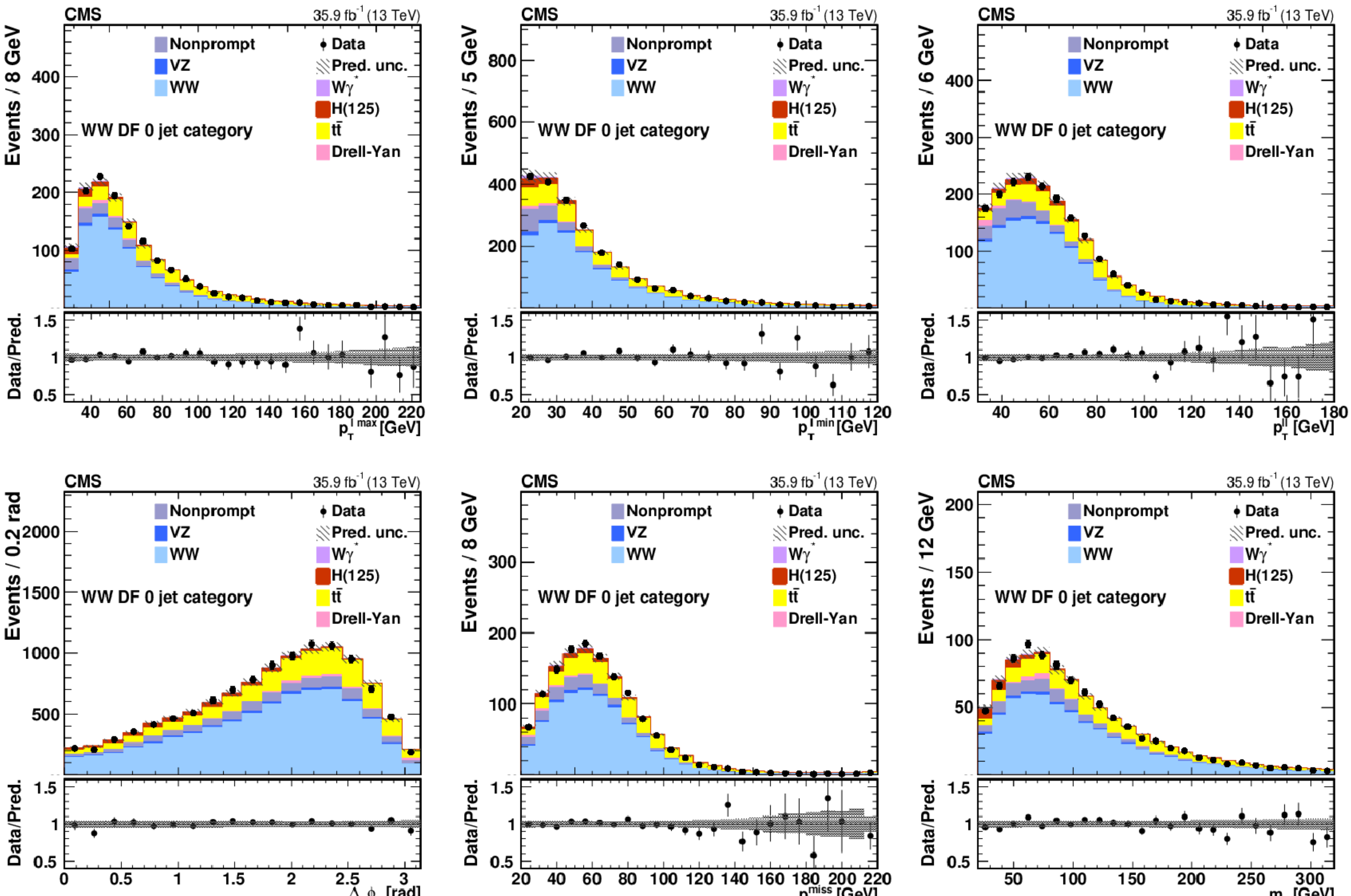
DYMVA: Developed for the CMS HWW analysis (arXiv:1806.05246)

*

$$projected\ TrkE_T^{miss} = \begin{cases} TrkE_T^{miss} & \Delta\phi_{\min}(leptons, TrkE_T^{miss}) \geq \pi/2 \\ TrkE_T^{miss}\sin\Delta\phi_{\min} & \Delta\phi_{\min}(leptons, TrkE_T^{miss}) \leq \pi/2 \end{cases}$$

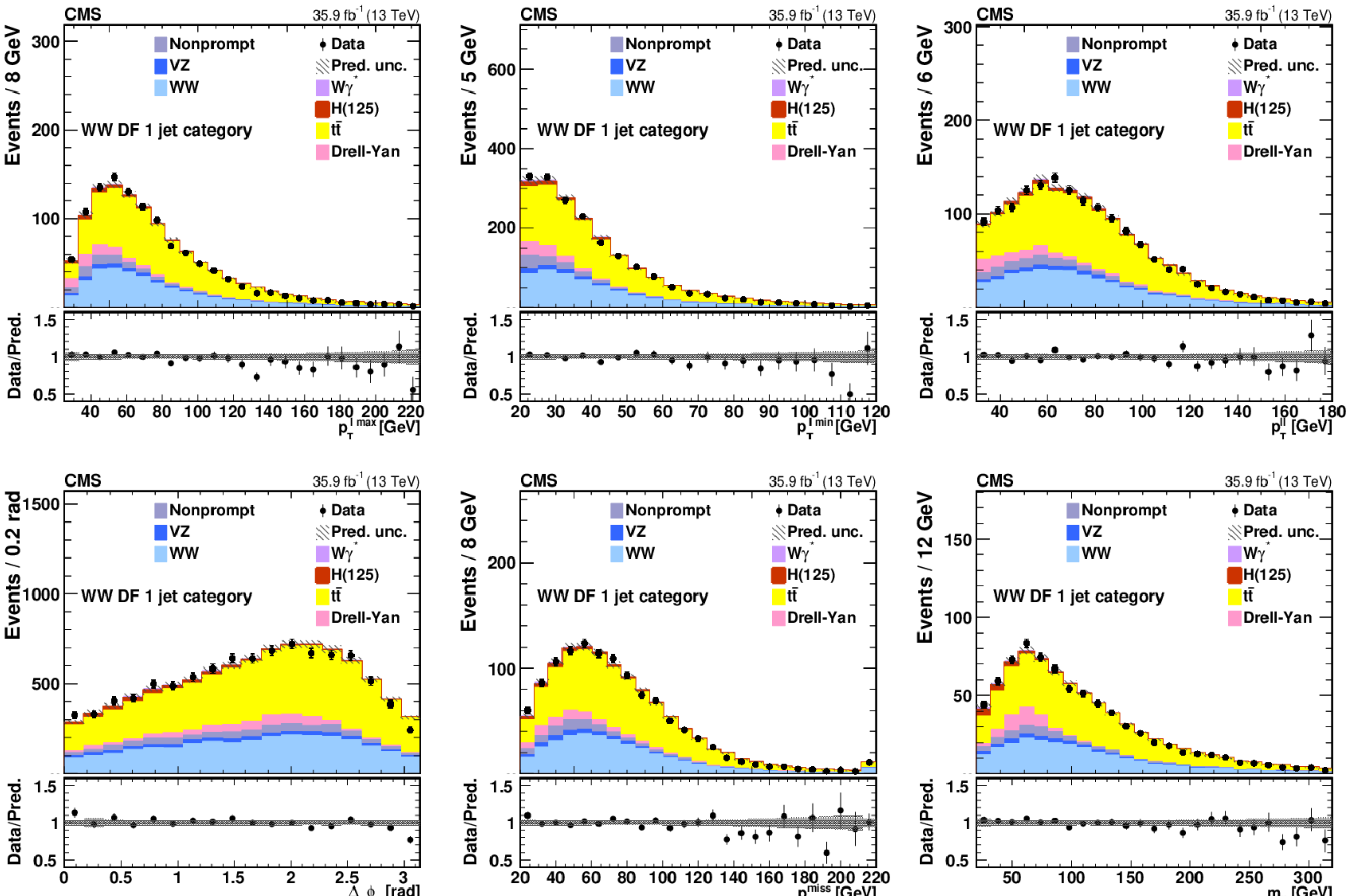
$$projected\ E_T^{miss} = \begin{cases} E_T^{miss} & \Delta\phi_{\min}(leptons, E_T^{miss}) \geq \pi/2 \\ E_T^{miss}\sin\Delta\phi_{\min} & \Delta\phi_{\min}(leptons, E_T^{miss}) \leq \pi/2 \end{cases}$$

Sequential analysis



1 jet category

Sequential analysis



Random Forest analysis

- Alternative approach: **Random Forest** (RF) multivariate analysis
  - Each individual tree is allowed to use only a random subset of variables. This approach **mitigates overfitting**
- **Pre-selection:** $m_{\ell\ell}$ > 30 GeV, third loose lepton veto (p_T > 10 GeV), bVeto (p_{Tj} > 20 GeV), $|m_{\ell\ell} - m_Z|$ > 15 GeV for same-flavour leptons
- After the preselection, the largest contamination comes from Drell-Yan and ttbar.
  Two independent RF have been trained
  - **anti-Drell-Yan:** WW vs DY
  - **anti-top:** WW vs ttbar
- Hyperparameters of the RFs are optimized by evaluating the RF performance in a multidimensional grid, taking into account all possible combinations between several choices for parameter values

Feature	Classifier	
	Drell–Yan	Top quark
Lepton flavor	✓	
Number of jets		✓
$p_T^{\ell\,\mathrm{min}}$	✓	
p_T^{miss}	✓	✓
$p_T^{\mathrm{miss,proj}}$	✓	
$p_T^{\ell\ell}$	✓	✓
$m_{\ell\ell}$	✓	
$m_{\ell\ell p_T^{\mathrm{miss}}}$	✓	
$\Delta\phi_{\ell\ell p_T^{\mathrm{miss}}}$	✓	✓
$\Delta\phi_{\ell J}$		✓
$\Delta\phi_{p_T^{\mathrm{miss}} J}$		✓
$\Delta\phi_{\ell\ell}$	✓	
H_T		✓
Recoil	✓	✓

RF optimized architecture:
ntrees= 50, max_depth = 20
min_events_per_split = 50
min_events_in_leaf = 1
max_features_per_tree = sqrt(total_variables)

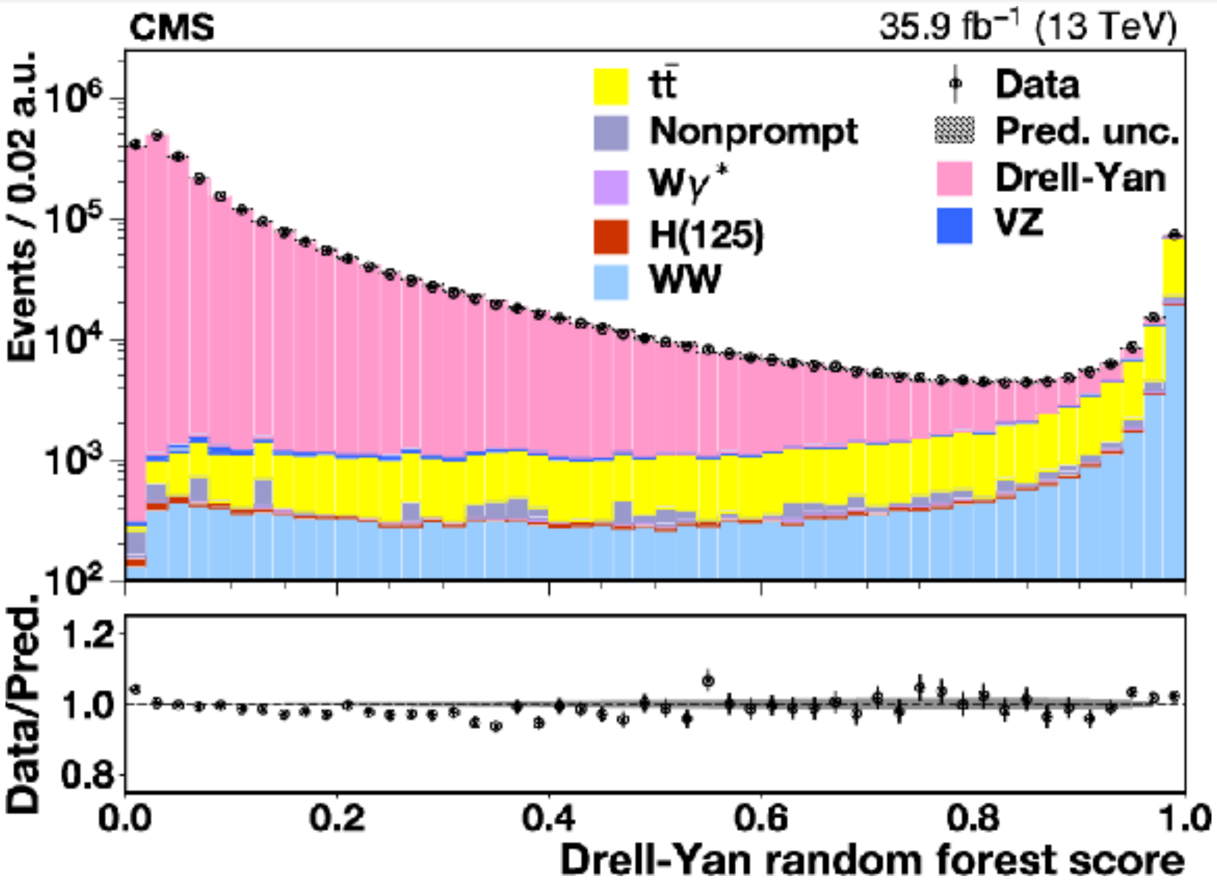
Random Forest analysis

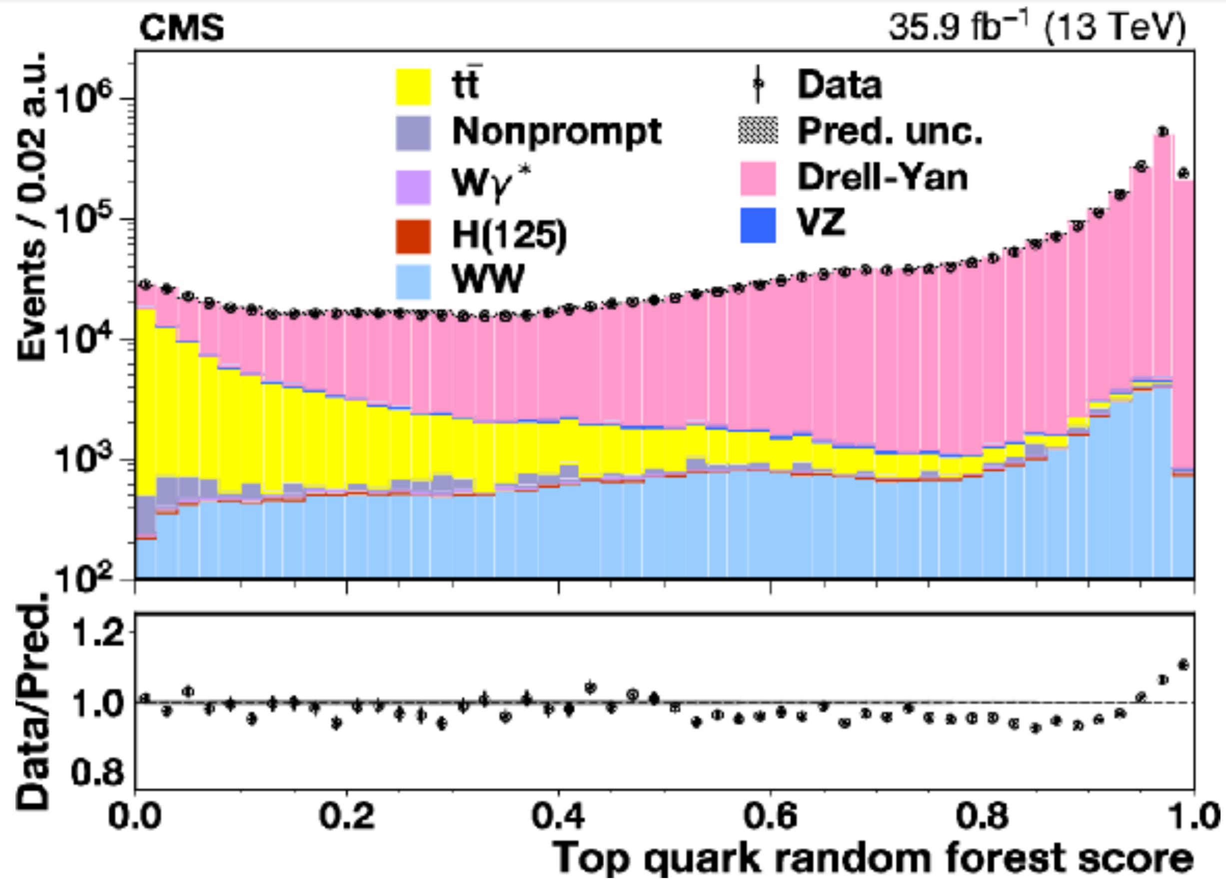
Signal Region definition:

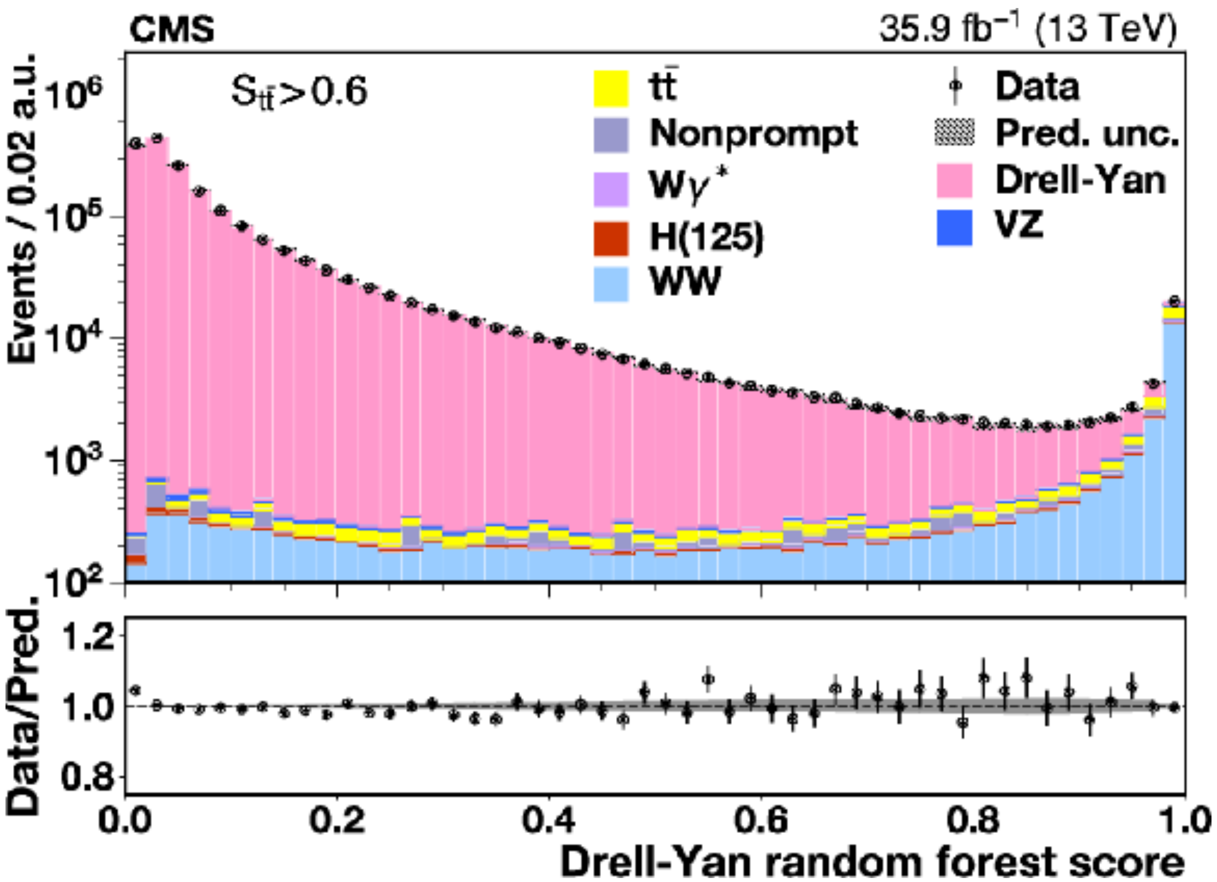
Quantity	Random Forest DF	Random Forest SF
Number of leptons	Strictly 2	
Lepton charges	Opposite	
$p_{\mathrm{T}}^{\ell\,\max}$	>25	
$p_{\mathrm{T}}^{\ell\,\min}$	>20	
$m_{\ell\ell}$	>30	>30
Additional leptons	0	
$\lvert m_{\ell\ell} - m_{\mathrm{Z}} \rvert$	—	>15
Number of b-tagged jets	0	
Drell–Yan RF score S_{DY}	>0.96	
$\mathrm{t\bar{t}}$ RF score $S_{\mathrm{t\bar{t}}}$	>0.6	

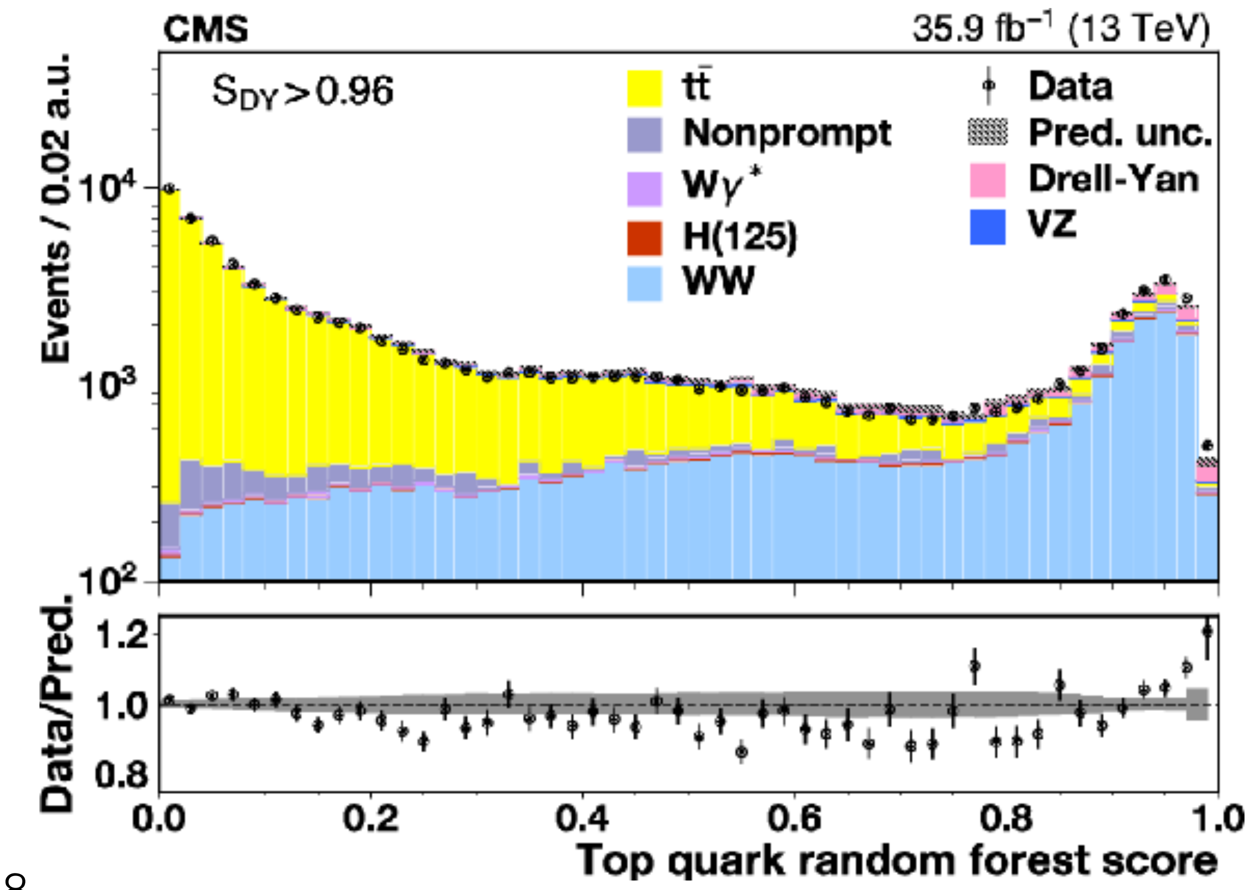
- Selections on RF scores have been optimized by simultaneously **minimizing the uncertainty** in the cross section and **maximizing the signal purity**

Random Forest analysis





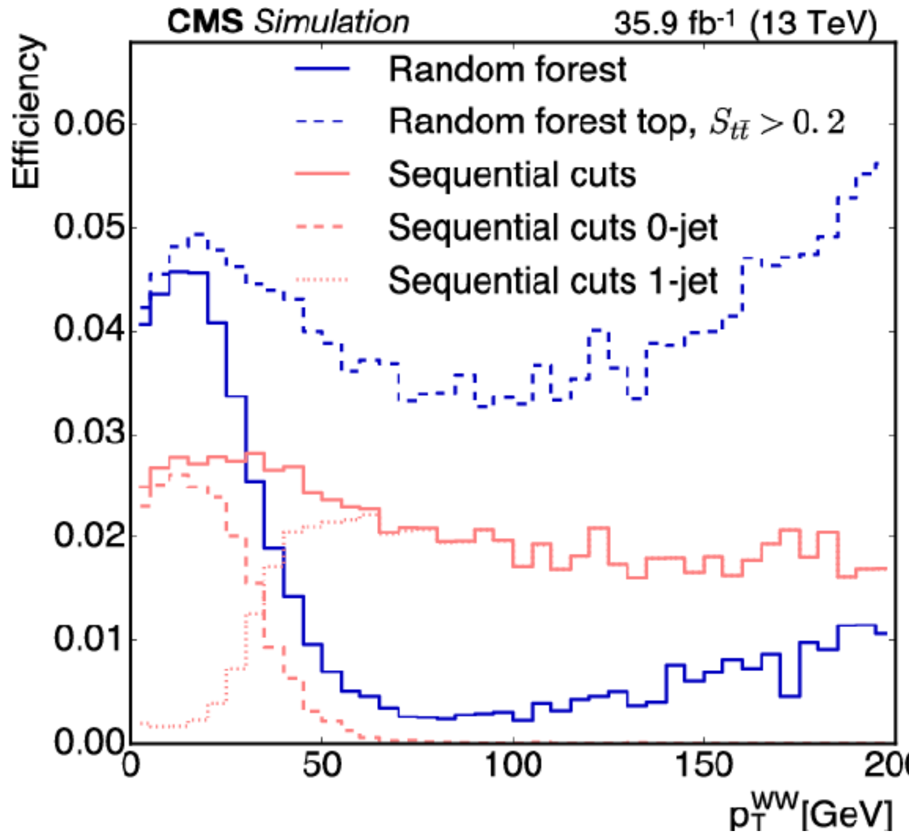




Total Xsec measurements

- In both approaches the signal strength is extracted by fitting the predicted yields to the observed events (1-bin distribution). Information from the control regions is included in the fit
    - **SEQ fit:** 4 Signal Regions, 4 Top Control Regions (2 flavour categories x 2 njets categories)
    - **RF fit:** 1 SR, 1 TopCR, 1 DYCR, 1 Same-SignedCR

Theoretical prediction: σ_{tot}^{NNLO} = 118.8 ± 3.6 pb



- ## Sequential analysis result:

Category		Signal strength	Cross section [pb]
0-jet	DF	1.054 ± 0.083	125.2 ± 9.9
0-jet	SF	1.01 ± 0.16	120 ± 19
1-jet	DF	0.93 ± 0.12	110 ± 15
1-jet	SF	0.76 ± 0.20	89 ± 24
0-jet & 1-jet	DF	1.027 ± 0.071	122.0 ± 8.4
0-jet & 1-jet	SF	0.89 ± 0.16	106 ± 19
0-jet & 1-jet	DF & SF	0.990 ± 0.057	117.6 ± 6.8

σ_{tot}^{SEQ} = 117.6 ± 1.4 (stat) ± 5.5 (syst) ± 1.9 (theo) ± 3.2 (lumi) pb = **117.6 ± 6.8 pb**

- ## Random Forest analysis result:

σ_{tot}^{RF} = 131.4 ± 1.3 (stat) ± 6.0 (syst) ± 5.1 (theo) ± 3.5 (lumi) pb = **131.4 ± 8.7 pb**

- Random forest gets a purer signal region. However, its sensitivity is concentrated at low pTWW due to jet-multiplicity related variables used in the training → more sensitive to theoretical uncertainties of pTWW spectrum corrections than the sequential analysis

Fiducial Xsec measurement

- **Fiducial region definition at gen level:** two dressed electrons or muons in the event with p_T > 20 GeV and |η|< 2.5, $m_{\ell\ell}$ > 20 GeV, $p_{T\ell\ell}$ > 30 GeV and E_T^{Miss} > 20 GeV

- **Results:** (Different-Flavour + Same-Flavour combination)

Theoretical prediction: σ_{fid}^{NNLO} = 1.531 ± 0.043 pb

$\boldsymbol{\sigma_{fid}^{tot}}$ = 1.529 ± 0.0020 (stat) ± 0.069 (syst) ± 0.028 (theo) ± 0.041 (lumi) pb = **1.529 ± 0.087 pb**

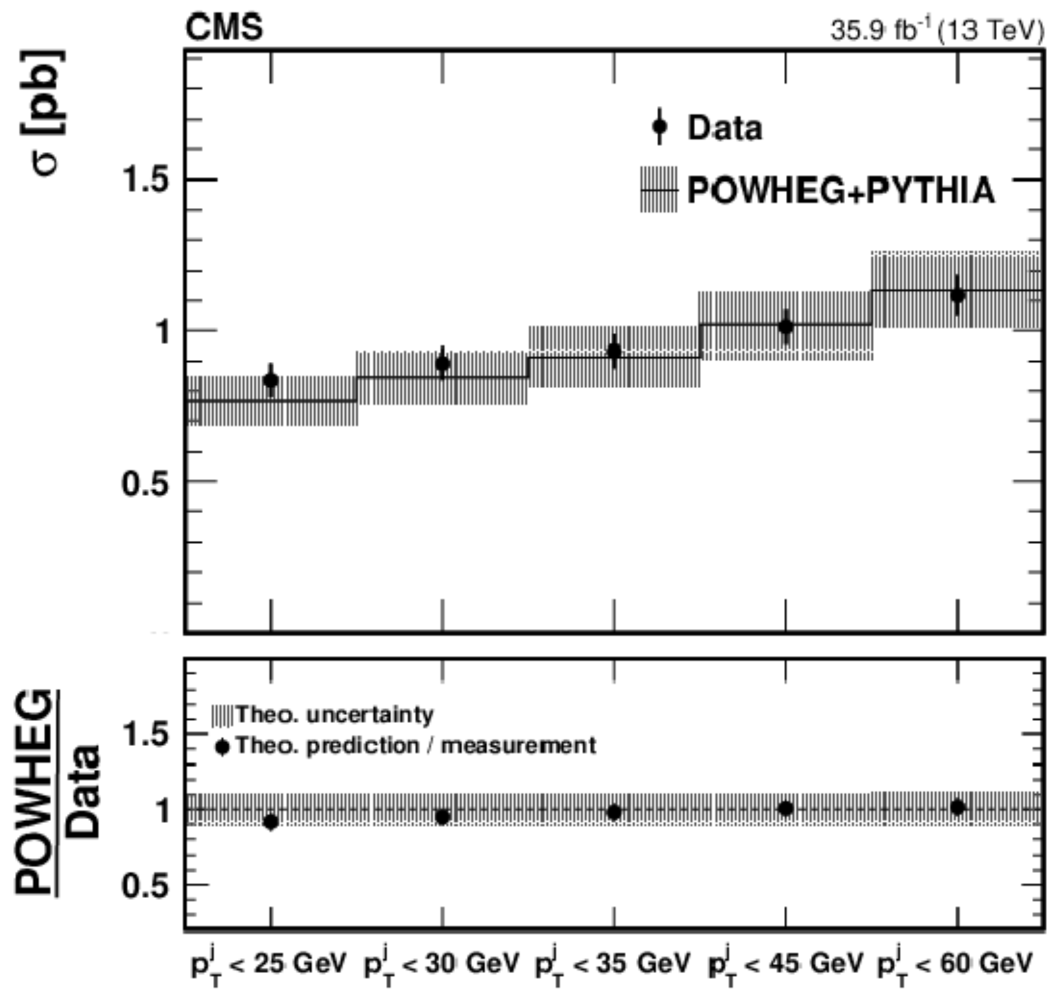
σ_{fid}^{tot} (based on 0 reco jets subset only) = 1.61 ± 0.10 pb

σ_{fid}^{tot} (based on 1 reco jets subset only) = 1.35 ± 0.11 pb

Fiducial WW+0 AK4 gen jets, pT thres. varied

p_T threshold (GeV)	Signal strength	Cross section (pb)
25	1.091 ± 0.073	0.836 ± 0.056
30	1.054 ± 0.065	0.892 ± 0.055
35	1.020 ± 0.060	0.932 ± 0.055
45	0.993 ± 0.057	1.011 ± 0.058
60	0.985 ± 0.059	1.118 ± 0.067

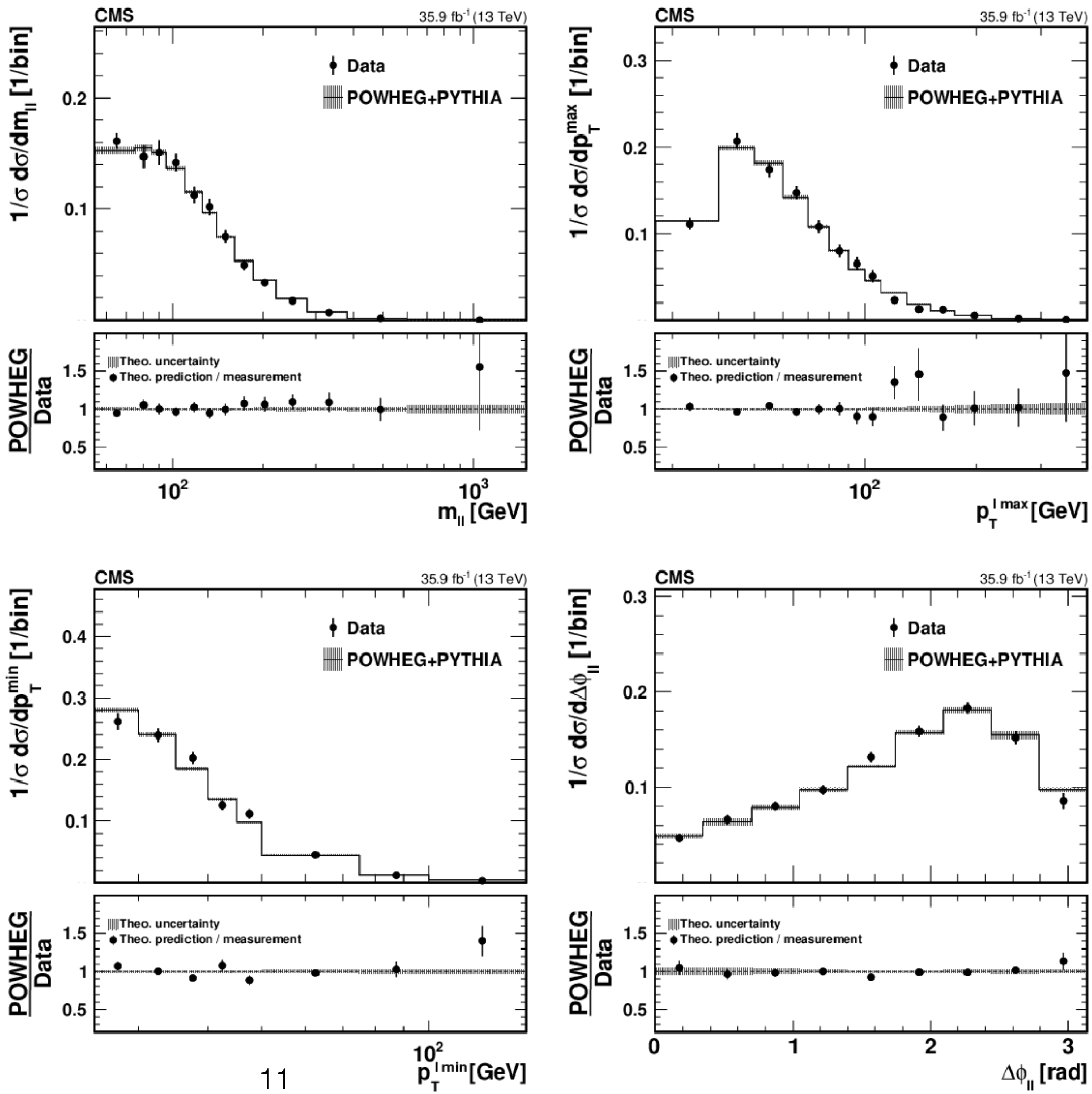
Fid WW+0 AK4 gen jets jet pT thres. varied



SEQ

Normalized differential Xsecs

- Differential cross section measurement in $m_{\ell\ell}$, $p_{T\ell1}$, $p_{T\ell2}$, $\Delta\phi_{\ell\ell}$ bins
- Using the same fiducial definition
- **Approach:** several signal strengths (bins categorized at GEN level) are fitted in RECO bins
    - The simultaneous fit to all bins in a given histogram takes all the correlations into account

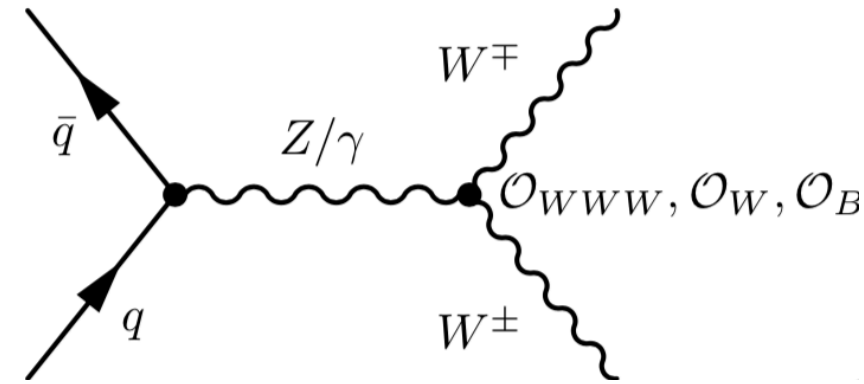


SEQ

Limits on Wilson coefficients

- In the electroweak sector of the SM, the first higher-dimensional operators containing only massive boson fields are dimension-6

  - Set limits on the 3 corresponding coefficients affecting WW production: **EFT effects simulated with Madgraph5 aMC@NLO**

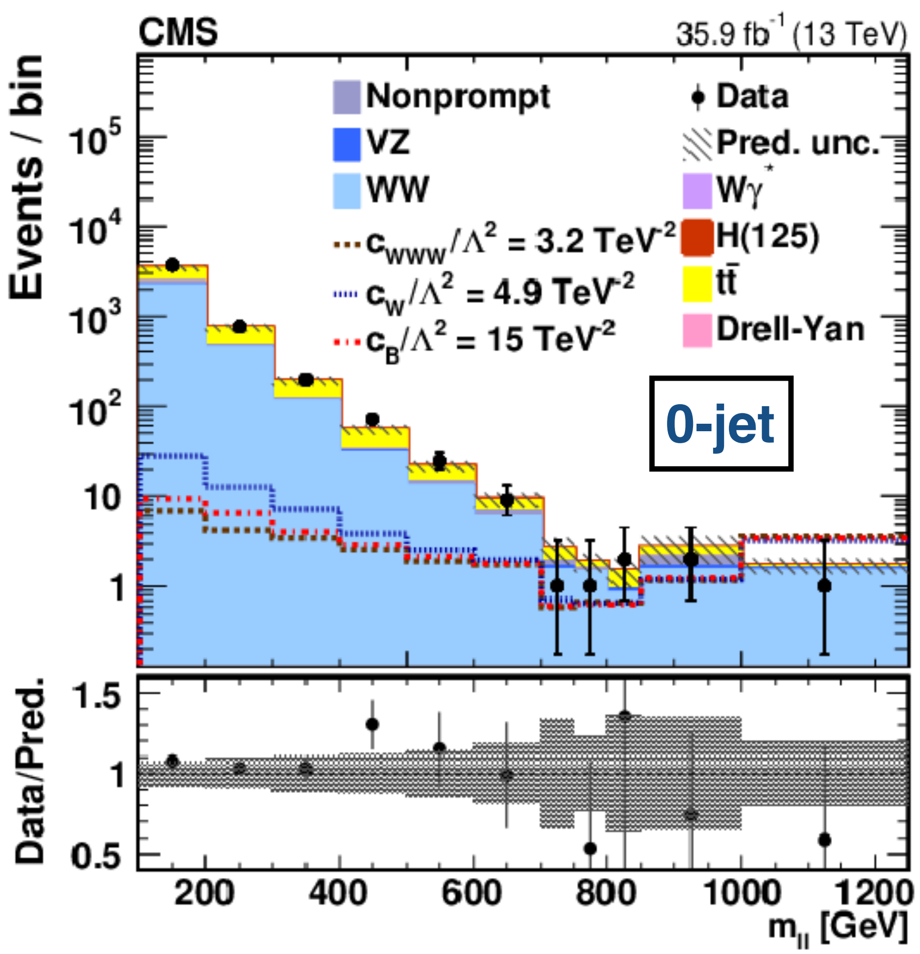


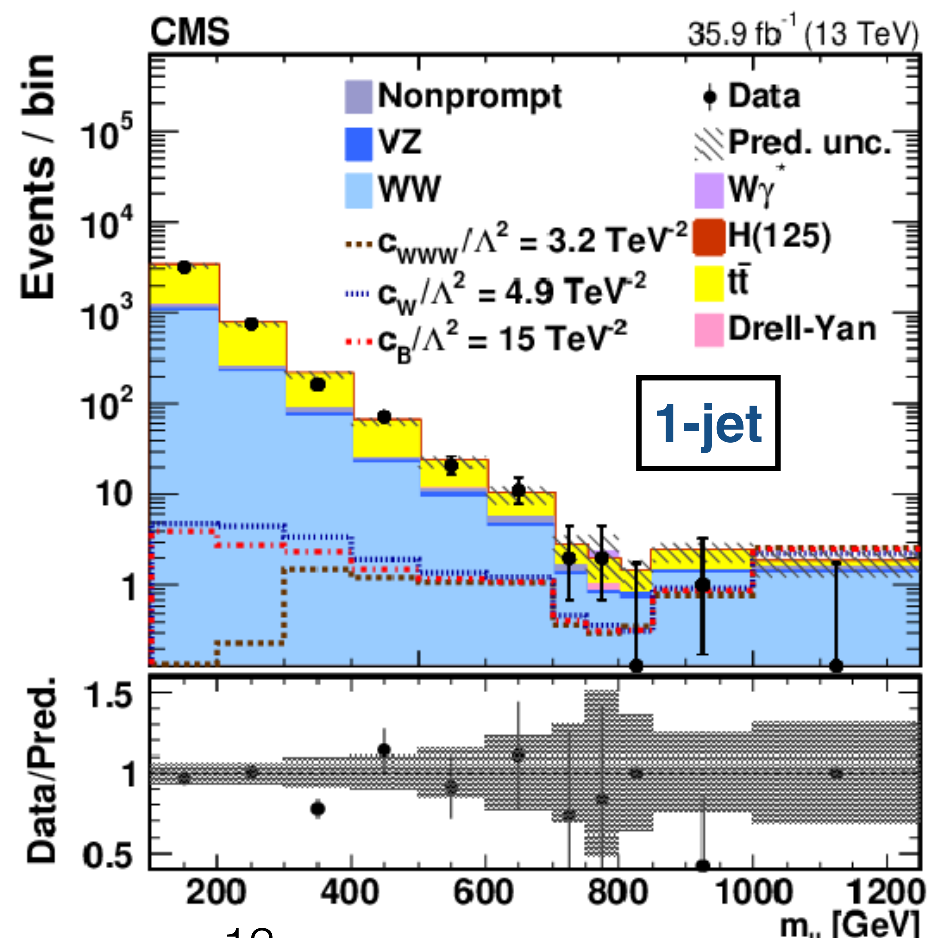
$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_{\rho}^{\ \mu},$$

$$\mathcal{O}_{W} = \frac{c_{W}}{\Lambda^2} (D^{\mu}\Phi)^{\dagger} W_{\mu\nu} (D^{\nu}\Phi),$$

$$\mathcal{O}_{B} = \frac{c_{B}}{\Lambda^2} (D^{\mu}\Phi)^{\dagger} B_{\mu\nu} (D^{\nu}\Phi),$$

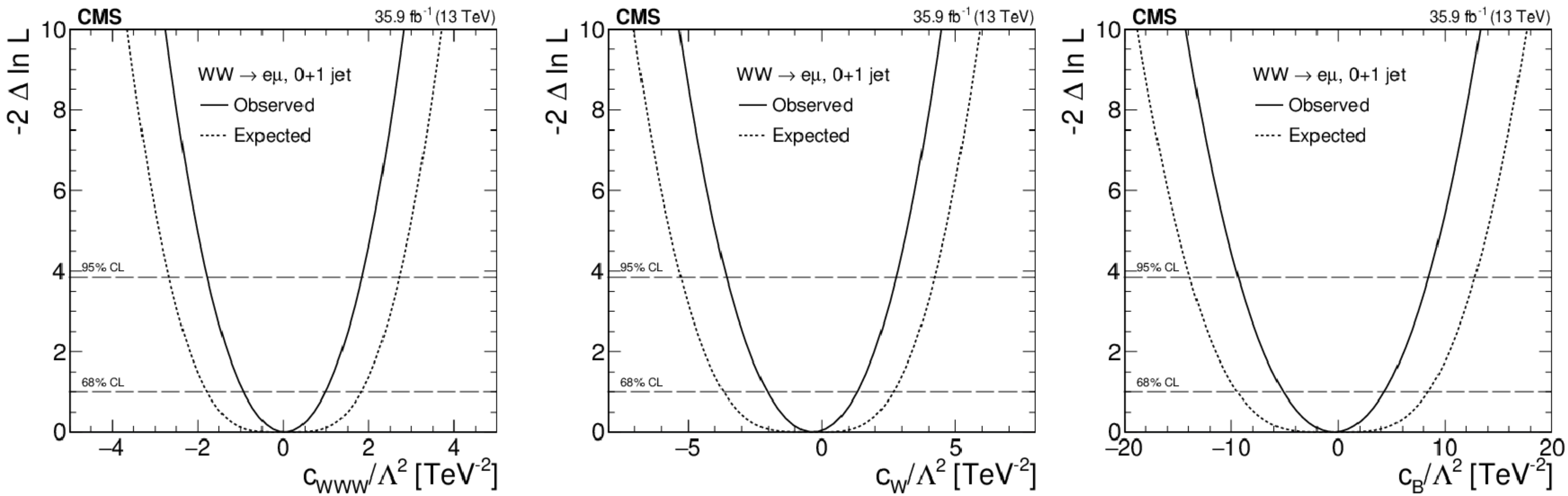
Used eμ final state from the sequential analysis





SEQ

Limits on Wilson coefficients

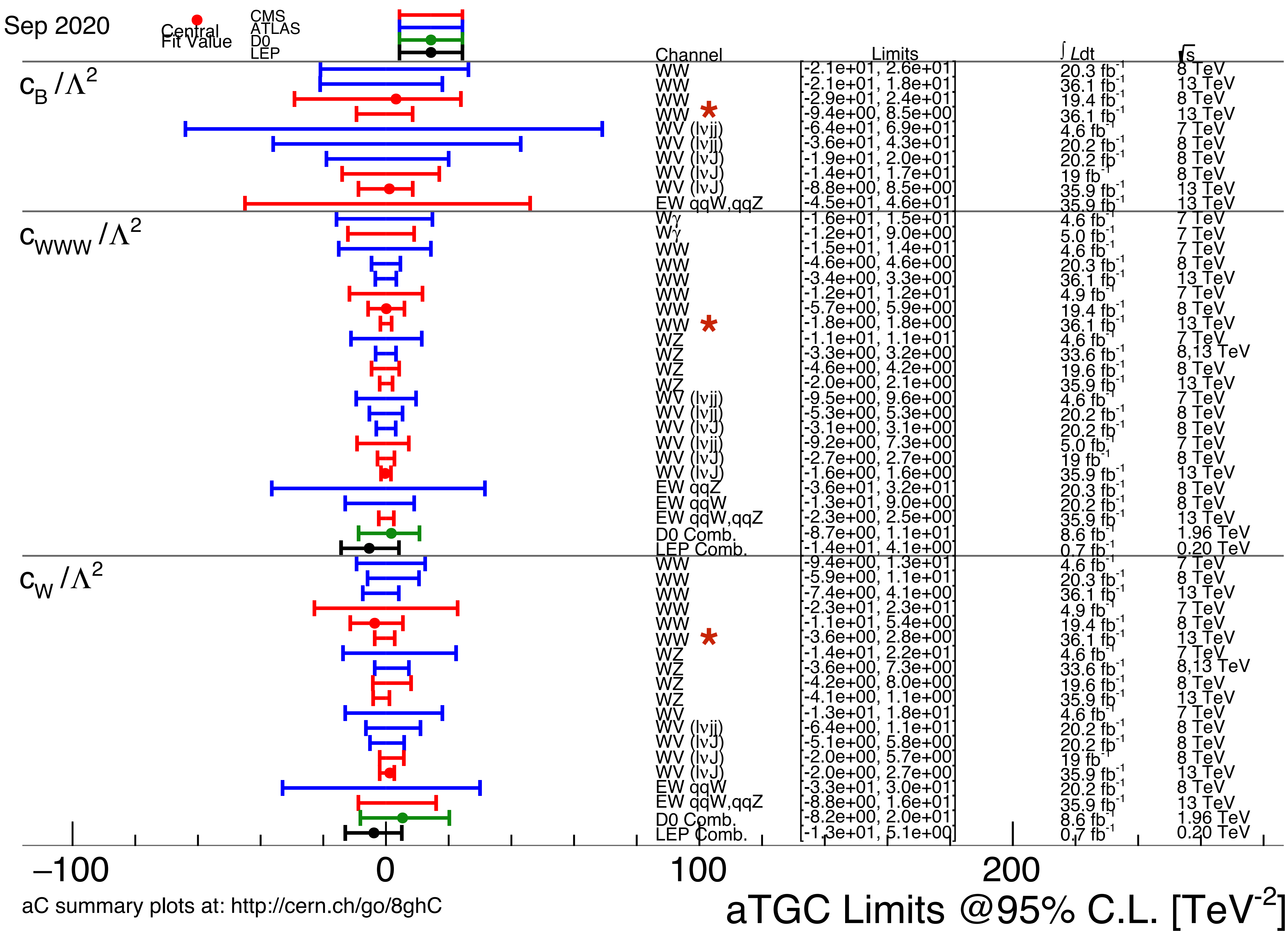


Coefficients (TeV^{-2})	68% confidence interval expected	68% confidence interval observed	95% confidence interval expected	95% confidence interval observed
c_{WWW}/Λ^2	$[-1.8, 1.8]$	$[-0.93, 0.99]$	$[-2.7, 2.7]$	$[-1.8, 1.8]$
c_W/Λ^2	$[-3.7, 2.7]$	$[-2.0, 1.3]$	$[-5.3, 4.2]$	$[-3.6, 2.8]$
c_B/Λ^2	$[-9.4, 8.4]$	$[-5.1, 4.3]$	$[-14, 13]$	$[-9.4, 8.5]$

W.r.t. Run-I (Observed): arXiv:1507.03268

Coupling constant	This result (TeV^{-2})	Its 95% CL interval (TeV^{-2})
c_{WWW}/Λ^2	$0.1^{+3.2}_{-3.2}$	$[-5.7, 5.9]$
c_W/Λ^2	$-3.6^{+5.0}_{-4.5}$	$[-11.4, 5.4]$
c_B/Λ^2	$-3.2^{+15.0}_{-14.5}$	$[-29.2, 23.9]$

Limits on Wilson coefficients



Summary & plans

- The WW production at 13 TeV results using the CMS full 2016 dataset experiment have been shown, including:
    - Total WW cross-section measurement
    - Fiducial & differential cross-section measurements
    - Limits on Wilson coefficients

- **Future plans of the analysis:**
    - Differential + aTGCs analysis using the full Run2 dataset (2016+2017+2018 CMS data)

Backup

Postfit yields

Process	Sequential Cut				Random Forest	
	DF		SF		DF	SF
	0-jet	1-jet	0-jet	1-jet	all jet multiplicities	
Top quark	2110 ± 110	5000 ± 120	1202 ± 66	2211 ± 69	3450 ± 340	830 ± 82
Drell–Yan	129 ± 10	498 ± 38	1230 ± 260	285 ± 86	1360 ± 130	692 ± 72
VZ	227 ± 13	270 ± 12	192 ± 12	110 ± 7	279 ± 29	139 ± 10
V V V	11 ± 1	29 ± 2	4 ± 1	6 ± 1	13 ± 4	3 ± 2
$H \to W^+W^-$	269 ± 41	150 ± 25	50 ± 2	27 ± 1	241 ± 26	90 ± 10
$W\gamma^{(*)}$	147 ± 17	136 ± 13	123 ± 5	58 ± 6	305 ± 88	20 ± 6
Nonprompt leptons	980 ± 230	550 ± 120	153 ± 39	127 ± 32	940 ± 300	183 ± 59
Total background	3870 ± 260	6640 ± 180	2950 ± 270	2820 ± 120		
	$10\,510 \pm 310$		5780 ± 300		6600 ± 480	1960 ± 120
$q\bar{q} \to W^+W^-$	6430 ± 250	2530 ± 140	2500 ± 180	1018 ± 71	$12\,070 \pm 770$	2820 ± 180
$gg \to W^+W^-$	521 ± 66	291 ± 38	228 ± 32	117 ± 15	693 ± 44	276 ± 17
Total W^+W^-	6950 ± 260	2820 ± 150	2730 ± 190	1136 ± 72		
	9780 ± 300		3860 ± 200		$12\,770 \pm 820$	3100 ± 200
Total yield	$10\,820 \pm 360$	9460 ± 240	5680 ± 330	3960 ± 360		
	$20\,280 \pm 430$		9640 ± 490		$19\,360 \pm 950$	5060 ± 240
Purity	0.64	0.30	0.48	0.29		
	0.48		0.40		0.66	0.61
Observed	10 866	9404	5690	3914	19 418	5210

Rel. syst. uncertainties in total Xsec

SEQ analysis: combined SRs + TopCRs

Uncertainty source	(%)
Statistical	1.2
$t\bar{t}$ normalization	2.0
Drell–Yan normalization	1.4
$W\gamma^*$ normalization	0.4
Nonprompt leptons normalization	1.9
Lepton efficiencies	2.1
b tagging (b/c)	0.4
Mistag rate (q/g)	1.0
Jet energy scale and resolution	2.3
Pileup	0.4
Simulation and data control regions sample size	1.0
Total experimental systematic	4.6
QCD factorization and renormalization scales	0.4
Higher-order QCD corrections and p_T^{WW} distribution	1.4
PDF and α_S	0.4
Underlying event modeling	0.5
Total theoretical systematic	1.6
Integrated luminosity	2.7
Total	5.7

RF only

Jet multiplicity measurement

- Relaxed cut on Stt (Stt > 0.2) to increase the efficiency for WW events with jets

Efficiency for RF selection w.r.t. preselection

Number of jets	0	1	≥ 2
Efficiency	0.555 ± 0.003	0.448 ± 0.004	0.290 ± 0.004

$$\vec{v} = R_{PU} R_{DET} \vec{t}$$
$$\vec{u} = R_{PU}^{-1} R_{DET}^{-1} \vec{v}$$
$$\vec{\omega} = \vec{u} / |\vec{u}|$$

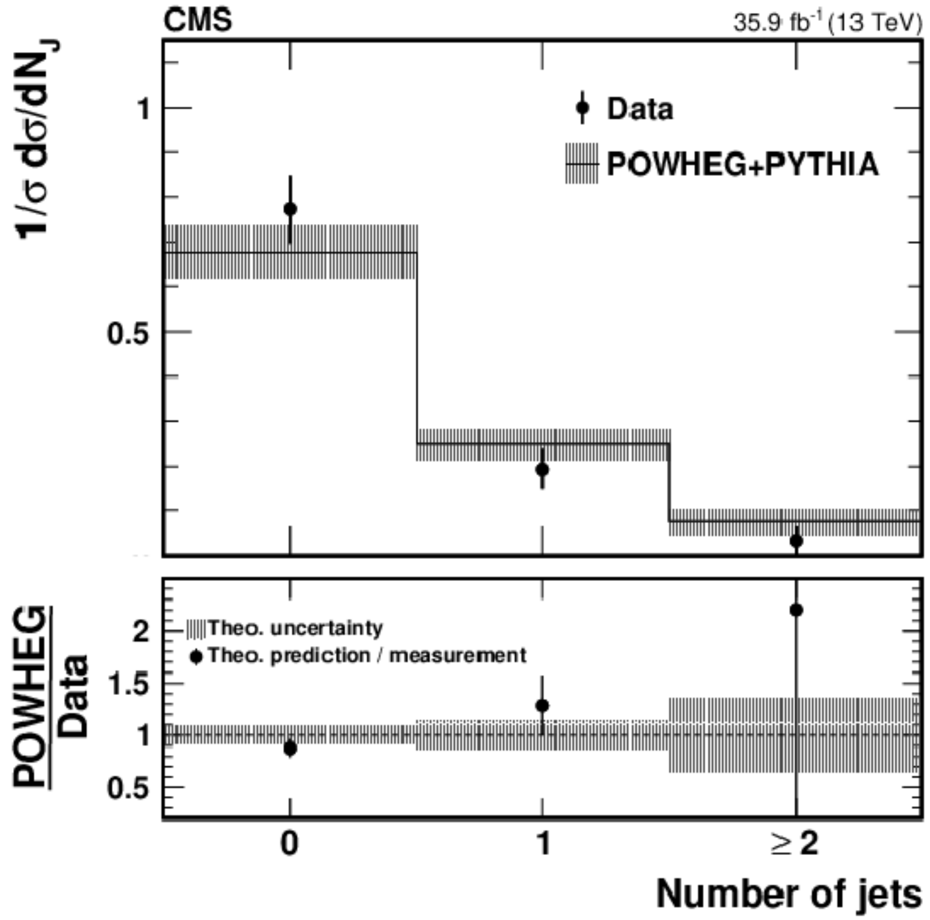
- **Unfolding:** Gen jets reconstructed from stable gen particles excluding neutrinos with p_Tj > 30 GeV and $|\eta| < 2.4$, separated from leptons by $\Delta R > 0.4$
  - Reconstructed and generated jets are said to match if $\Delta R_{gen,reco} < 0.4$

$$\mathbf{R}_{\mathrm{PU}} = \begin{pmatrix} 0.986 & 0 & 0 \\ 0.013 & 0.985 & 0 \\ 0.001 & 0.015 & 1 \end{pmatrix}$$

$$\mathbf{R}_{\mathrm{DET}} = \begin{pmatrix} 0.963 & 0.060 & 0.003 \\ 0.036 & 0.891 & 0.090 \\ 0.001 & 0.049 & 0.906 \end{pmatrix}$$

Total relative uncert.

$$\begin{pmatrix} 0.011 & 0.193 & 0.374 \\ 0.210 & 0.007 & 0.140 \\ 0.305 & 0.181 & 0.015 \end{pmatrix}$$

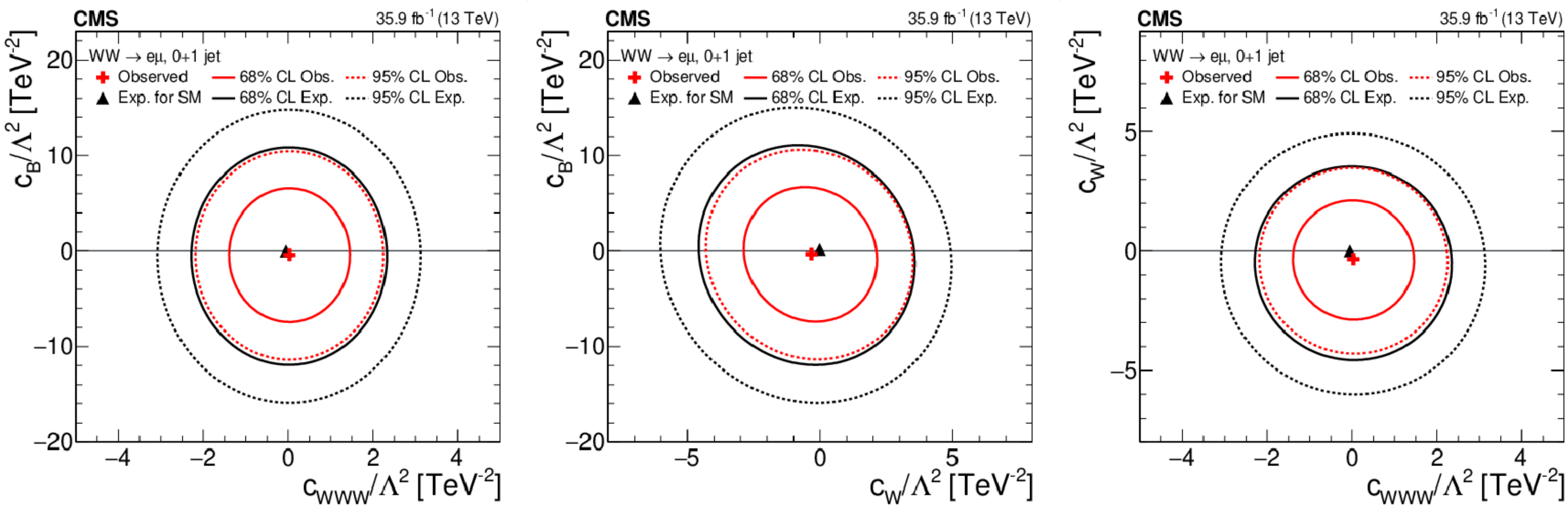


Fraction of events

Number of jets	0	1	≥ 2
Before unfolding	$0.795 \pm 0.007 \pm 0.053$	$0.180 \pm 0.006 \pm 0.039$	$0.025 \pm 0.005 \pm 0.018$
After unfolding	$0.773 \pm 0.008 \pm 0.075$	$0.193 \pm 0.007 \pm 0.043$	$0.034 \pm 0.006 \pm 0.033$
Predicted	$0.677 \pm 0.007 \pm 0.058$	$0.248 \pm 0.007 \pm 0.033$	$0.075 \pm 0.006 \pm 0.026$

SEQ

Limits on Wilson coefficients



W.r.t. Run-I results: arXiv:1507.03268