Lessons from VBF *W*/*Z* from Run 1

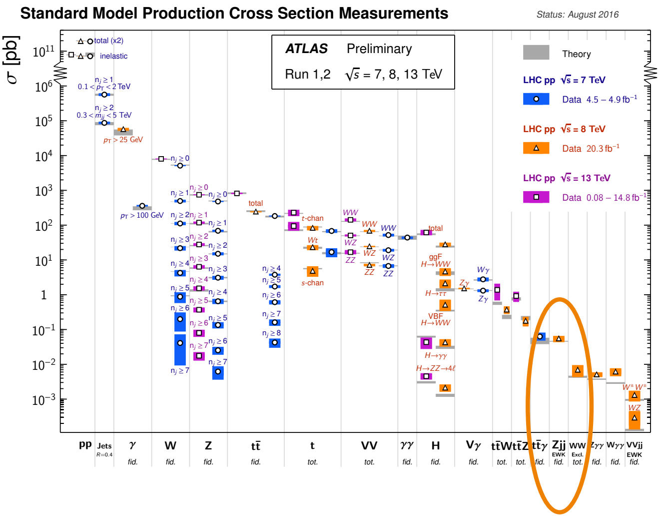
Christian Gütschow

Future of VBF measurements workshop, IPPP

21 September 2016

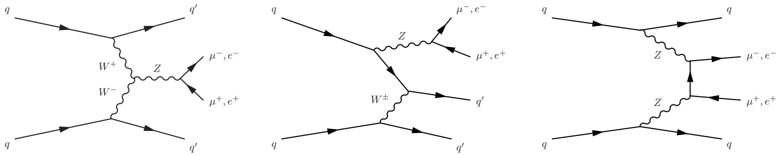
Overview

- CMS measurement of Zjj at 7 TeV: JHEP 10 (2013) 062
- ATLAS measurement of Zjj at 8 TeV: JHEP 04 (2014) 031
- CMS measurement of Zjj at 8 TeV: EPJC 75 (2015) 66
- CMS measurement of Wjj at 13 TeV: arXiv:1607.06975, submitted to JHEP
- ATLAS measurement of Wjj at 13 TeV: soon



Signal definition

- electroweak Vjj production: gauge invariant set of diagrams with t-channel exchange of a heavy gauge boson



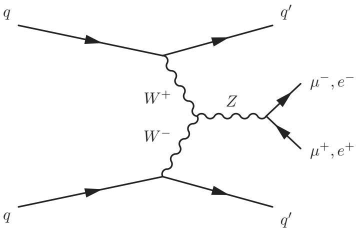
- negative interference between the VBF diagram and V bremsstrahlung diagrams

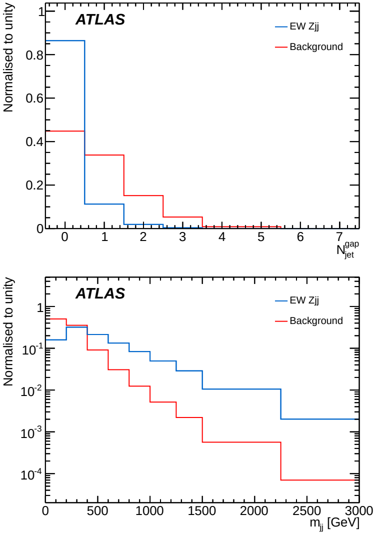
Theoretical Predictions

- ATLAS
  - fully simulated QCD $Zjj + 0, 1, 2j$@LO baseline sample using `Sherpa v1.4.3` (`CT10`) with `MENLOPS`
  - particle-level QCD $Zjj + 0j$@NLO+$1j$@LO prediction using `Powheg` (`CT10`) + `Pythia6` (`Perugia2011`)
  - fully simulated EW $Zjj + 0, 1j$@LO baseline sample using `Sherpa v1.4.3` (`CT10`)
  - particle-level EW $Zjj + 0j$@LO prediction using `Powheg` (`CT10`) + `Pythia6` (`Perugia2011`)
- CMS
  - fully simulated QCD $Zjj + 0, 1, 2j$@LO baseline sample using `Madgraph` (`CTEQ6L1`) + `Pythia6` (`Z2*`)
  - fully simulated EW $Zjj + 0j$@LO baseline sample using `Madgraph` (`CTEQ6L1`) + `Pythia6` (`Z2*`)
  - fully simulated QCD $Wjj + 0, 1, 2j$@LO baseline sample using `MG5_aMC@NLO v2.1` (`CTEQ6L1`) + `Pythia6` (`Z2*`)
  - fully simulated EW $Wjj + 0j$@LO baseline sample using `MG5_aMC@NLO v2.1` (`CTEQ6L1`) + `Pythia6` (`Z2*`)

Main signal features

- colourless exchange of vector bosons
- jets tend to be produced at large rapidities with sizeable transverse momentum and together they balance the gauge boson



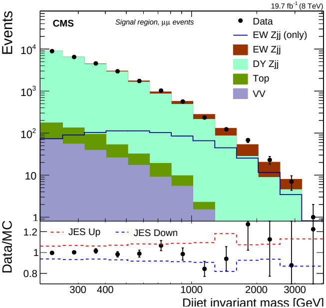


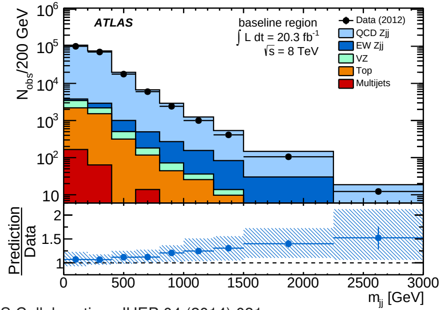
ATLAS Collaboration, JHEP 04 (2014) 031

- results in **two forward jets with large dijet invariant mass** and **no additional jets in the rapidity interval** between them

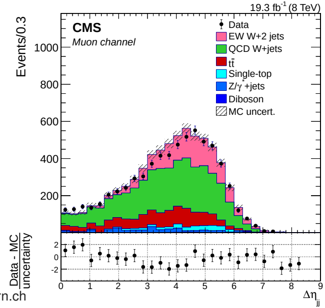
Backgrounds

CMS Collaboration, EPJC 75 (2015) 66





ATLAS Collaboration, JHEP 04 (2014) 031



CMS Collaboration, arXiv:1607.06975

Zjj@ATLAS strategy

- select events requiring a Z candidate with tight invariant mass cut and at least two high-p_T jets
- use jet multiplicity in rapidity interval between leading two jets to construct signal-enhanced ($N_{\text{jet}}^{\text{gap}} = 0$) and signal-suppressed ($N_{\text{jet}}^{\text{gap}} \geq 1$) regions
- use signal-suppressed control region to **constrain shape of the background model**
- extract electroweak *Zjj* component in signal-enhanced search region using two-component template **fit to dijet invariant mass spectrum**

region	yield for N_{EW}
$m_{jj} > 250$ GeV	1657^{+134}_{-132} (fit) ± 40 (MC)
$m_{jj} > 1$ TeV	333^{+27}_{-26} (fit) ± 8 (MC)

- convert signal yields into fiducial cross sections and **correct back to particle level** using a correction factor ($\mathcal{C}_{\text{EW}}$)

Zjj@CMS strategy

- select events requiring a *Z* candidate with moderate invariant mass cut and at least two 30 GeV jets (50 GeV for measurement)
- constrain background modelling using **simulation-based approach** and, alternatively, **data-driven approach**
- fit for electroweak *Zjj* component after training BDT and requiring $s/b > 10\,\%$
- convert signal yields into signal strength and fiducial cross sections

Wjj@CMS strategy

- select events requiring a *W* candidate with moderate transverse mass cut and at least two high-p_T jets
- cut tight on dijet invariant mass (1 TeV)
- train a BDT to separate out QCD *Wjj*, obtain QCD *Wjj* normalisaton scale factors (~ 0.7) from data after cutting on BDT output
- extract electroweak *Wjj* component using unbinned maximum-likelihood fit to m_{jj} spectrum
- modell m_{jj} shape with $\left(m_{jj}^{a_0 + a_1 \ln(m_{jj}/8000)}\right)^{-1}$ where a_0 and a_1 are obtained from simulation

Zjj@ATLAS selection

object	search	control
leptons	$\lvert\eta^{\ell}\rvert < 2.47,\ p_{\mathrm{T}}^{\ell} > 25\,\mathrm{GeV}$	
dilepton pair	$81\,\mathrm{GeV} \leq m_{\ell\ell} \leq 101\,\mathrm{GeV}$ $p_{\mathrm{T}}^{\ell\ell} > 20\,\mathrm{GeV}$	
jets	$\lvert y^{j}\rvert < 4.4,\ \Delta R_{j,\ell} \geq 0.3$ $p_{\mathrm{T}}^{j_1} > 55\,\mathrm{GeV},\ p_{\mathrm{T}}^{j_2} > 45\,\mathrm{GeV}$	
dijet system	$m_{jj} > 250\,\mathrm{GeV}$	
interval jets	$N_{\mathrm{jets}}^{\mathrm{gap}} = 0$	$N_{\mathrm{jets}}^{\mathrm{gap}} \geq 1$
Zjj system	$p_{\mathrm{T}}^{\mathrm{balance}} < 0.15$	$p_{\mathrm{T}}^{\mathrm{balance,3}} < 0.15$

where

$$(\Delta R_{j,\ell})^2 = (\Delta\phi_{j,\ell})^2 + (\Delta\eta_{j,\ell})^2$$

and

$$p_{\mathrm{T}}^{\mathrm{balance}} = \frac{\left|\vec{p}_{\mathrm{T}}^{\,\ell_1} + \vec{p}_{\mathrm{T}}^{\,\ell_2} + \vec{p}_{\mathrm{T}}^{\,j_1} + \vec{p}_{\mathrm{T}}^{\,j_2}\right|}{\left|\vec{p}_{\mathrm{T}}^{\,\ell_1}\right| + \left|\vec{p}_{\mathrm{T}}^{\,\ell_2}\right| + \left|\vec{p}_{\mathrm{T}}^{\,j_1}\right| + \left|\vec{p}_{\mathrm{T}}^{\,j_2}\right|}$$

Zjj@CMS selection

- BDT trained using dijet and and Z boson kinematic variables that are robust against jet energy scale, **simulation based**
- BDT trained using full event information in muon channel only, **simulation based**
- BDT trained using only dijet kinematic variables since background constrained with **$\gamma+$ jets**

Analysis	*A*	*B*	*C*
Channels	ee, $\mu\mu$	$\mu\mu$	ee, $\mu\mu$ binned in M_{jj}
Selection	$p_{Tj1,j2} > 50, 30$ GeV		
	$R p_T^{hard} < 0.14$, $\lvert y^* \rvert < 1.2$, $M_{jj} > 200$ GeV		$p_{TZ} > 50$ GeV, $\lvert yZ \rvert < 1.4442$, $M_{jj} > 450$ GeV
Jets	PF	JPT	PF
Variables used			
M_{jj}	•	•	•
p_{Tj_1}, p_{Tj_2}		•	•
η_{j_1}, η_{j_2}			•
$\Delta_{rel}(jj) = \frac{\lvert \vec{p}_{Tj_1} + \vec{p}_{Tj_2} \rvert}{p_{Tj_1} + p_{Tj_2}}$			•
$\Delta\eta_{jj}$		•	
$\lvert \eta_{j_1} \rvert + \lvert \eta_{j_2} \rvert$	•	•	•
$\Delta\phi_{jj}$		•	•
$\Delta\phi_{Z,j_1}$		•	
yZ	•	•	
$z^* Z$	•		
p_{TZ}	•	•	
$R p_T^{hard}$		•	
q/g discriminator	•		•
DY Zjj model	MC-based	MC-based	From data

CMS Collaboration, EPJC 75 (2015) 66

Wjj@CMS selection

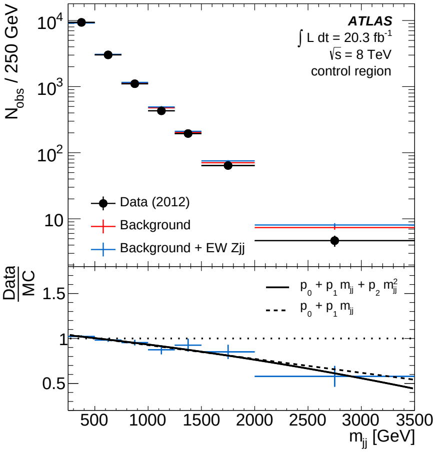
$W \to \ell\nu$ Lepton requirements	Jet requirements
Single lepton trigger	$p_{T}{}^{j1} > 60\,\text{GeV}$, $p_{T}{}^{j2} > 50\,\text{GeV}$
High-quality lepton ID and isolation	$\lvert y_{W} - (y_{j1} + y_{j2})/2 \rvert < 1.2$
Electron (muon) $p_{T} > 30$ (25) GeV	$m_{jj} > 1000\,\text{GeV}$
$E_{T}^{miss} > 30$ (25) GeV for electron (muon) channels	
W transverse mass $> 30\,\text{GeV}$	
Veto second lepton	

CMS Collaboration, arXiv:1607.06975

Background modelling constraint in ATLAS

- fit of data-to-MC ratio in **control region used to constrain shape of background model in search region**
- get strong *Zjj* modelling directly from data
- **constrains systematics** on background model
- theory uncertainties $> 30\,\%$ and experimental uncertainties $> 20\,\%$ if background shape not constrained

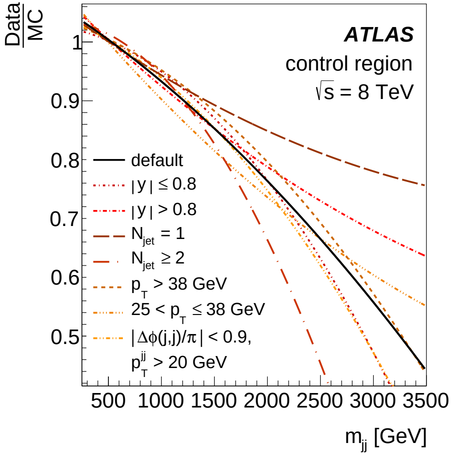
dijet invariant mass in control region



ATLAS Collaboration, JHEP 04 (2014) 031

Choice of control region

- nominal control region split into 6 different subregions probing additional jet activity in rapidity interval between leading jets:
  - nominal control region split based on rapidity of in-gap jet
  - nominal control region split based on p_{T} of in-gap jet
  - nominal control region split based on number of in-gap jets
- MPI-suppressed subregion:
  - nominal control region with dijet $p_{\mathrm{T}}^{jj} > 20\,\mathrm{GeV}$ and $|\Delta\phi(j,j)| < 0.9\pi$

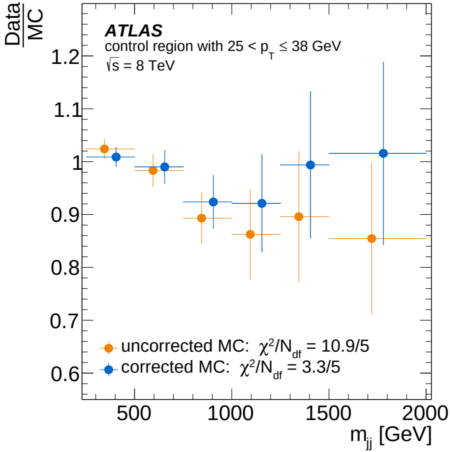


ATLAS Collaboration, JHEP 04 (2014) 031

- **consistent signal yield with maximum 5 % spread** between subregions, i.e. within statistical uncertainty of extrapolation ($\sim 10\,\%$)

Proof of principle

- extrapolation procedure examined by correcting subregions with constraints derived in complementary subregion, e.g.
  - split nominal control region based on p_T of in-gap jet
  - correct one subregion with constraint derived in the other subregion
- in all cases, **corrected simulation describes data better than uncorrected simulation**



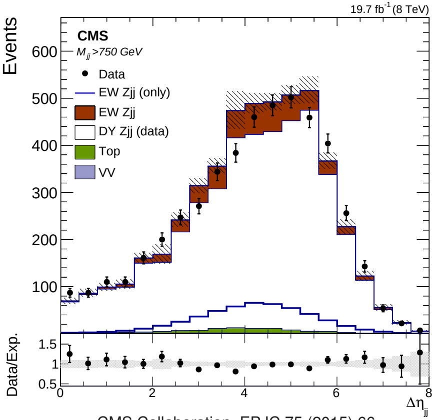
ATLAS Collaboration, JHEP 04 (2014) 031

Background modelling constraint in CMS

- **simulation-based approach**: reweight LO Madgraph sample on event-by-event basis dynamically ($=$ as a function of m_{jj}) using parton-level MCFM calculations at NLO and LO

- **data-driven approach**: tighten boson kinematics and reweight $p_T(\gamma)$ to $p_T(Z)$, assume dijet kinematics are the same for $m_{jj} > 2m_Z$

- check data-driven method using γjj simulation and derive systematic



CMS Collaboration, EPJC 75 (2015) 66

Interference effects

- so far all experiments estimate the interference using a 3-sample strategy:
  - generate a LO sample with both QCD *Vjj* and EW *Vjj* diagrams at matrix-element level
  - generate a LO sample with only QCD *Vjj* diagrams at matrix-element level
  - generate a LO sample with only EW *Vjj* diagrams at matrix-element level
  - estimate interference contribution by subtracting QCD-only and EW-only from combined sample
- interference contribution tends to be one of the dominant uncertainties
- How can we do better in future measurements?

Zjj@ATLAS systematic uncertainties

Source	ΔN_{EW}		$\Delta \mathcal{C}_{EW}$	
	Electrons	Muons	Electrons	Muons
Lepton systematics	—	—	±3.2 %	±2.5 %
Control region statistics	±8.9 %	±11.2 %	—	—
Jet-energy scale	±5.6 %		$^{+2.7}_{-3.4}$ %	
Jet-energy resolution	±0.4 %		±0.8 %	
Pileup jet modelling	±0.3 %		±0.3 %	
Jet-vertex fraction	±1.1 %		$^{+0.4}_{-1.0}$ %	
Signal modelling	±8.9 %		$^{+0.6}_{-1.0}$ %	
Background modelling	±7.5 %		—	
Signal/background interference	±6.2 %		—	
Parton distribution function	$^{+1.5}_{-3.9}$ %		±0.1 %	

EXPERIMENTAL

THEORY

Zjj@CMS systematic uncertainties

	Source	Shape	Methods A,B	Method C
Experimental	Luminosity	○	2.6	
	Trigger/selection	○	2–3	
	JES and residual jet response	●	1–10	
	JER	●	6–15	
	Pileup	●	6	
	Simulation statistics	●	variable	
	DY Zjj distribution (data)	●	—	variable
Theoretical	PDF	●	variable	
	μ_R/μ_F (signal)	●	variable	
	DY Zjj shape (MC)	●	variable	—
	DY Zjj shape (PDF and EW γjj contribution)	●	—	variable
	Interference	●	100	
	Normalisation of top-quark and diboson processes	○	7–10	

	Analysis A			Analysis B	Analysis C		
	ee	$\mu\mu$	ee + $\mu\mu$	$\mu\mu$	ee	$\mu\mu$	ee + $\mu\mu$
Luminosity	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Trigger/lepton selection	0.04	0.04	0.04	0.04	0.04	0.04	0.04
JES+residual response	0.06	0.05	0.05	0.04	0.06	0.05	0.05
JER	0.02	0.02	0.02	0.02	0.04	0.04	0.03
Pileup	0.01	0.02	0.02	0.01	0.01	0.01	0.01
DY Zjj	0.07	0.05	0.07	0.08	0.14	0.12	0.13
q/g discriminator	<0.01	<0.01	<0.01	—	<0.01	<0.01	<0.01
Top, dibosons	0.01	0.01	0.01	0.01	<0.01	<0.01	<0.01
Signal acceptance	0.03	0.04	0.04	0.04	0.06	0.06	0.06
DY/EW Zjj interference	0.14	0.14	0.14	0.13	0.06	0.08	0.08
Systematic uncertainty	0.19	0.18	0.19	0.17	0.17	0.17	0.18
Statistical uncertainty	0.11	0.10	0.07	0.09	0.24	0.21	0.16
$\mu = \sigma/\sigma_{\text{th}}$	0.82	0.86	0.84	0.89	0.91	0.85	0.88

CMS Collaboration, EPJC 75 (2015) 66

Wjj@CMS systematic uncertainties

Source of uncertainty	Electrons	Muons
Integrated luminosity	2.6%	
Jet energy scale	5.4%	7.3%
Jet energy resolution	2.2%	3.7%
QCD W+jets shape and normalization	13.0%	16.7%
Top quark background shape and normalization	5.5%	6.0%
Interference effect	14.4%	13.8%
Jets faking electrons fraction (electron channel)	4.4%	—
Lepton trigger efficiency	0.9%	1.0%
Lepton selection efficiency	1.8%	2.0%
Pileup	<1%	<1%
Fiducial acceptance	1.7%	1.7%
Total (without integrated luminosity)	21.6%	24.1%

CMS Collaboration, arXiv:1607.06975

Zjj@ATLAS search region ($m_{jj} > 250$ GeV)

➜ Measurement:

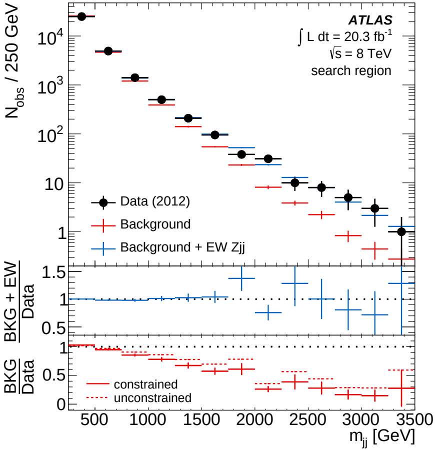
54.7 $\left\{ \begin{array}{ll} \pm 4.6 & \text{(stats)} \\ ^{+9.8}_{-10.4} & \text{(syst)} \\ \pm 1.5 & \text{(lumi)} \end{array} \right.$ **fb**

➜ Standard Model prediction:

46.1 $\left\{ \begin{array}{ll} \pm 0.2 & \text{(stats)} \\ ^{+0.3}_{-0.2} & \text{(scale)} \\ \pm 0.8 & \text{(PDF)} \\ \pm 0.5 & \text{(model)} \end{array} \right.$ **fb**

(using `PowhegPythia6`)

dijet invariant mass in search region



ATLAS Collaboration, JHEP 04 (2014) 031

Zjj@ATLAS search region with $m_{jj} > 1$ TeV

→ Measurement:

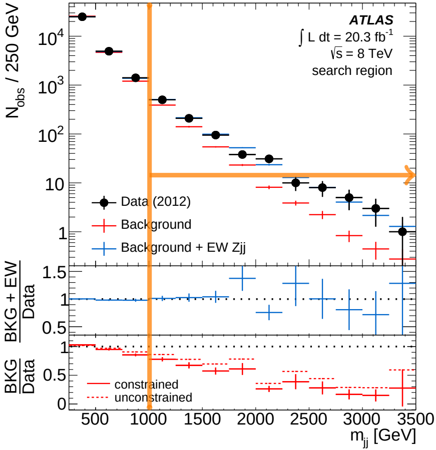
$$10.7 \begin{cases} \pm 0.9 & \text{(stats)} \\ \pm 1.9 & \text{(syst)} \\ \pm 0.3 & \text{(lumi)} \end{cases} \text{fb}$$

→ Standard Model prediction:

$$9.38 \begin{cases} \pm 0.05 & \text{(stats)} \\ ^{+0.15}_{-0.24} & \text{(scale)} \\ \pm 0.24 & \text{(PDF)} \\ \pm 0.09 & \text{(model)} \end{cases} \text{fb}$$

(using `PowhegPythia6`)

dijet invariant mass in search region



ATLAS Collaboration, JHEP 04 (2014) 031

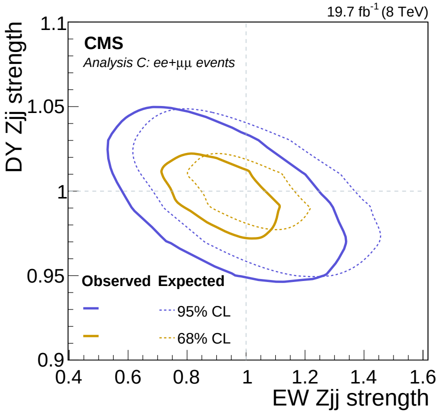
Wjj@CMS signal extraction

➜ Measurement:

$$174 \begin{cases} \pm 15 & \text{(stats)} \\ \pm 40 & \text{(syst)} \end{cases} \text{fb}$$

➜ Standard Model prediction:

$$208 \pm 18 \ \text{fb}$$



CMS Collaboration, EPJC 75 (2015) 66

Wjj@CMS signal extraction

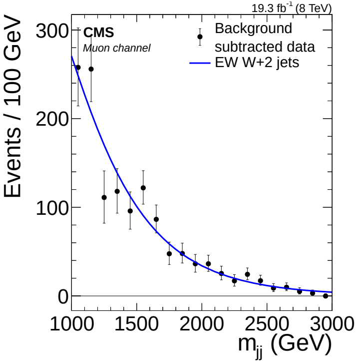
Channel	Measured cross section
Electron	$0.41 \pm 0.04\,(\text{stat}) \pm 0.09\,(\text{syst}) \pm 0.01\,(\text{lumi})\,\text{pb}$
Muon	$0.43 \pm 0.04\,(\text{stat}) \pm 0.10\,(\text{syst}) \pm 0.01\,(\text{lumi})\,\text{pb}$
Combined	$0.42 \pm 0.04\,(\text{stat}) \pm 0.09\,(\text{syst}) \pm 0.01\,(\text{lumi})\,\text{pb}$

CMS Collaboration, arXiv:1607.06975

➜ Standard Model prediction:

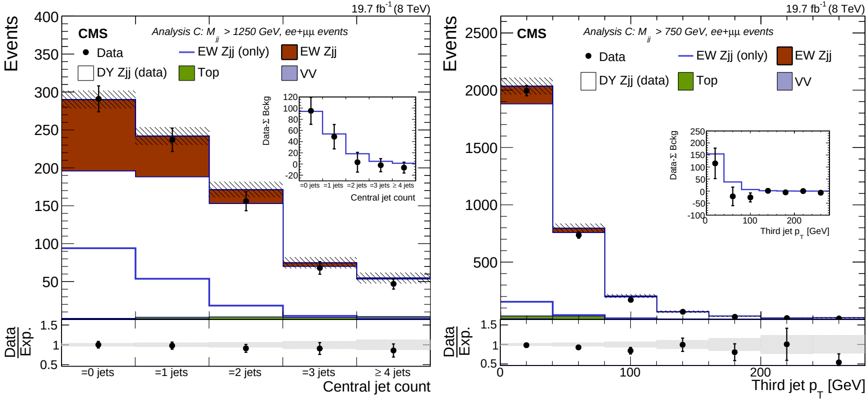
$$0.50 \begin{cases} \pm 0.02 & \text{(scale)} \\ \pm 0.02 & \text{(PDF)} \end{cases} \text{pb}$$

(using MG5_aMC@NLO+Pythia6)



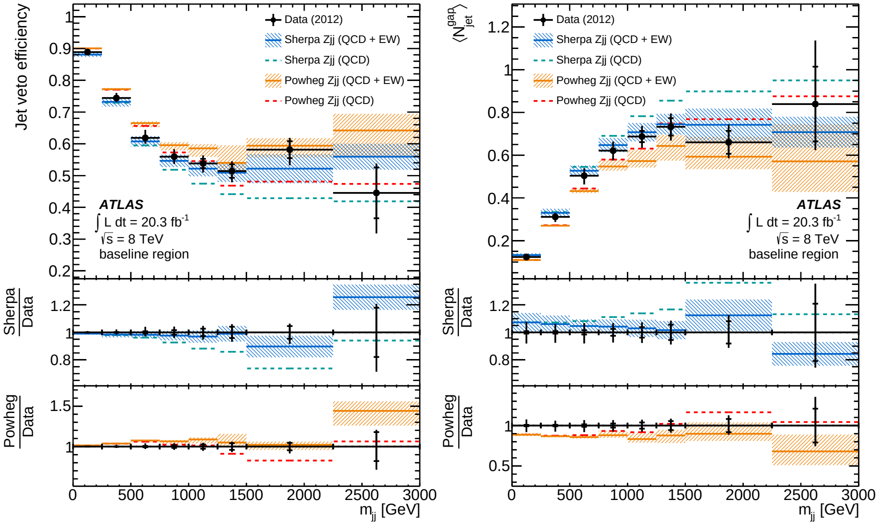
CMS Collaboration, arXiv:1607.06975

Zjj@CMS study of detector-level gap activity



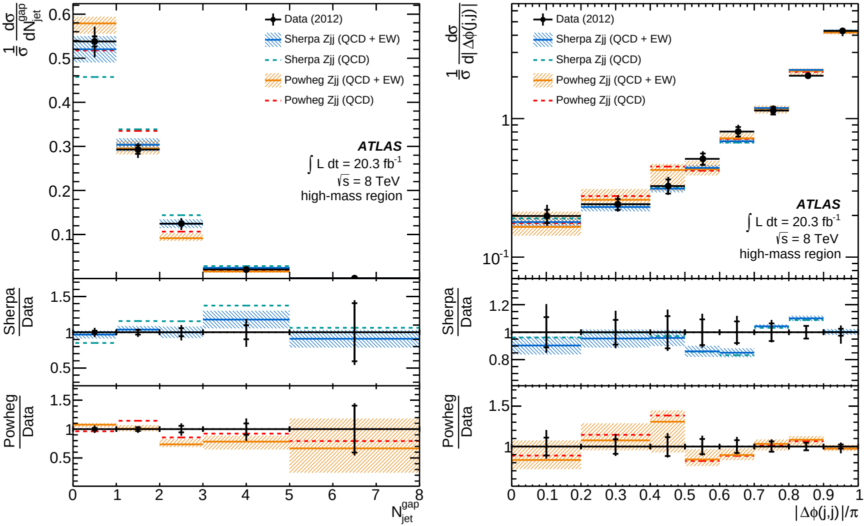
CMS Collaboration, EPJC 75 (2015) 66

Zjj@ATLAS study of particle-level gap activity



ATLAS Collaboration, JHEP 04 (2014) 031

ATLAS measurement of inclusive *Zjj* production



ATLAS Collaboration, JHEP 04 (2014) 031

Summary

- both EW *Zjj* and EW *Wjj* signals have been extracted and integrated cross sections have been measured
- ATLAS uses a jet veto to enhance the signal and uses the orthogonal region to constrain the background modelling in a data-driven way
- for *Zjj*, CMS uses a BDT to enhance the signal and constrain the background modelling with γ*jj* data
- for *Wjj*, CMS uses a BDT to enhance the background, thereby constraining the background normalisation using data
- the electroweak signal is then extracted using a maximum likelihood fit
- interference between QCD and EW *Vjj* is estimated using LO Monte Carlo samples
  - one of the dominant uncertainties – how can this be improved?
- CMS provided a study of the gap activity in *Zjj* events at the detector level
- ALTAS provided differential cross section measurements of inclusive $Z + 2$ jet production, including jet-veto efficiencies
  - is it worth the effort?