Search for the EW production of a VW pair plus two jets in the semi-leptonic $\ell\nu jj$ channel with full Run-II data

Asar Ahmed

Prof. Md. Naimuddin

Department of Physics & Astrophysics
University of Delhi, Delhi, India (110007)

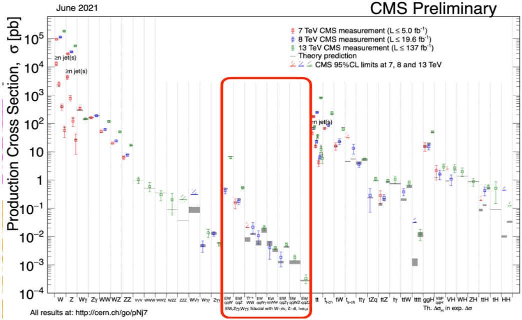
XXV DAE-BRNS High Energy Physics Symposium
IISER Mohali, Dec-2022

December 12, 2022

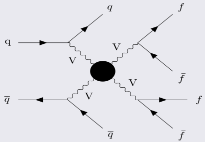
Vector Boson Scattering

So Far

- Tested and measured SM corss-sections over 10 orders of magnitude
- Discovered Higgs
- **Still need to understand the mechanism** behind the Electroweak Symmetry Breaking Mechanism (EWSB)



VBS Diagram



- Without Higgs, Vector Boson Scattering (VBS) cross-section violate unitarity at the TeV scale
- VBS: the radiated gauge bosons by quarks from the two protons and interact with each other and decay afterwards
- VBS measurement gives way to probe the EWSB
- Rare process in SM, requires good discrimination against enormous backgrounds
  - With advancement in ML techniques, provide way to improve signal discrimination
- First evidence of VBS in WV semi-leptonic channel at LHC.

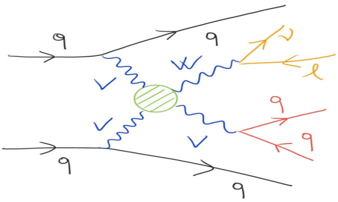
8.3%

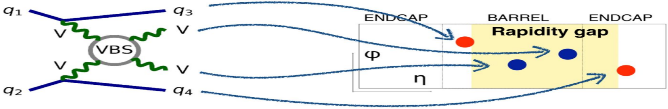
Signal signature

WV scattering is a EWK process with 6 fermions in the final state where V (W/Z) boson decays hadronically and the other W leptonically

VBS WV semileptonic channel not yet observed at LHC

- It has a good benefit of
  - **larger cross-section** than fully-leptonic one
- and drawback of
  - **larger background** from V+jets processes





VBS Topology

- High pseudo-rapidity gap between VBS jets: $\Delta\eta_{jj} > 2.5$
- Larger di-jet invariant mass: $M_{VBS-jj} > 500\ GeV$

16.7%

Objects and Samples

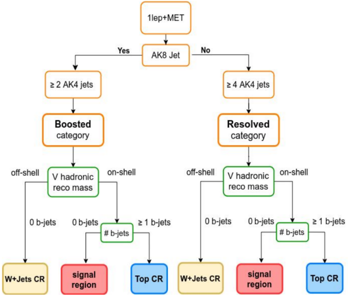
	2016	2017/2018
Electrons	EG POG mva_90p_Iso2016 ID (90% eff) + isolation	EG POG Fall17V1Iso_WP90 ID (90% eff) + isolation
Muons	Muon POG Tight selection + extra d_z d_{xy} cuts	Muon POG Tight selection + extra d_z d_{xy} cuts
Jets	AK4 Jets not overlapping with e/μ within ΔR of 0.4 + Tight Jet Id + Loose PU Jet ID p_T < 50 GeV	AK4 Jets not overlapping with e/μ within ΔR of 0.4 + Tight Jet Id + Loose PU Jet ID p_T < 50 GeV
FatJets	AK8PF jets $\lvert\eta\rvert < 2.4$, $p_T > 200$ GeV, ΔR > 0.8 from e/μ + Tight Jet Id + Loose PU Jet ID p_T < 50 GeV	AK8PF jets $\lvert\eta\rvert$ ¡ 2.4, $p_T > 200$ GeV, ΔR > 0.8 from e/μ + Tight Jet Id + Loose PU Jet ID p_T < 50 GeV (2017-only + Tight PU jet ID for η region $\lvert 2.5 - 3.2 \rvert$)
b-tag	BTV POG deepCSV: tight WP for b-tagging loose WP for b-veto	BTV POG deepCSV: tight WP for b-tagging loose WP for b-veto
MET	PuppiMET	PuppiMET

Signal and Background samples

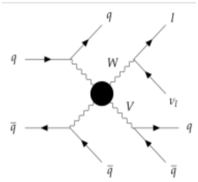
- **WVJJ (EWK) : Electroweak production of WVJJ**
- **WVJJ (QCD initiated): Irreducible background for analysis**
- **W+Jets: Most dominating background**
- **Drell-Yan: Z/γ decays to $\ell^+\ell^-$ and we mis-measure one lepton because of acceptance or ineffciency effects, gives missing energy**
- **tt Jets: Top quark always decays to one b-quark and one W boson. So, $t\bar{t} \rightarrow bWbW \rightarrow b\ell\nu\ell\nu$, if we mis-measure one lepton and b quark form jets**
- **Single top production: Here $t \rightarrow bW \rightarrow b\ell\nu$, and 3 jets is reconstructed**

25%

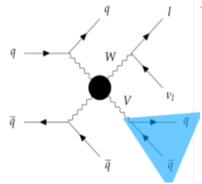
Analysis phase spaces



	Signal region	Top control region	W+jets control region
Resolved category	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet and Vjets p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bVeto with Loose** DeepCSV WP V had p_T < 200 GeV 65 GeV < Mjj Vhad < 105 GeV	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet and Vjets p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bTag with Tight** DeepCSV WP V had p_T < 200 GeV 65 GeV < Mjj Vhad < 105 GeV	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet and Vjets p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bVeto with Loose** DeepCSV WP V had p_T < 200 GeV 40 < Mjj Vhad < 65 GeV, Mjj Vhad > 105 GeV
Boosted category	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bVeto with Loose** DeepCSV WP V had p_T > 200 GeV 70 GeV < Mjj Vhad < 115 GeV	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bTag with Tight** DeepCSV WP V had p_T > 200 GeV 70 GeV < Mjj Vhad < 115 GeV	Ele p_T > 30 GeV (2016), 35 GeV (2017, 2018) Muon p_T > 30 GeV PuppiMET > 30 GeV Leading VBS jet p_T > 50 GeV trailing VBS jet p_T > 30 GeV $\Delta\eta_{VBS}$ > 2.5 , $M_{jj\ VBS}$ > 500 GeV Leptonic M^T_W < 185 GeV **bVeto with Loose** DeepCSV WP V had p_T > 200 GeV 40 GeV < Mjj Vhad < 70 GeV 115 GeV < Mjj Vhad < 250 GeV



Resolved category



Boosted category

VBS enhancement:

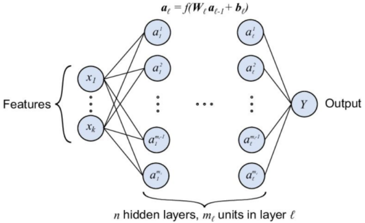
- Zeppendfield Variable: $z = \dfrac{\eta - \frac{\eta_{j1}+\eta_{j2}}{2}}{|\eta_{j1} - \eta_{j2}|}$
- Centrality: $\chi = \min(\Delta\eta_{-}, \Delta\eta_{+})$

33.3%

Deep Neural Network (DNN) for signal extraction

Approach

- All Backgrounds vs Signal (events weighted by XS and SFs)
- Training performed on events in signal region
- Trained 2 different models:
  - for boosted and resolved categories
  - joining all the years datasets
- Carefully checked and avoided overtraining



Resolved

- DNN fully connected (64-64-64-64 nodes)
- 16 inputs variables
- Regularization: Batch normalization and L2 norm weights normalization
- 1.7 M background, 220k signal samples

Boosted

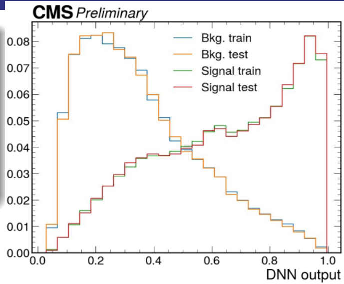
- DNN fully connected (64-32-32-32 nodes)
- 13 inputs variables
- Regularization: Batch normalization and L2 norm weights normalization
- 416k background, 50k signal samples

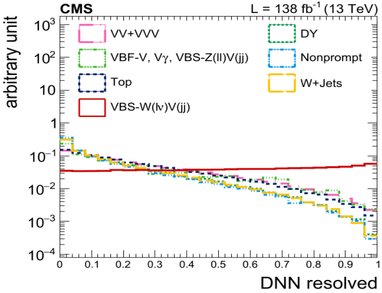
41.7%

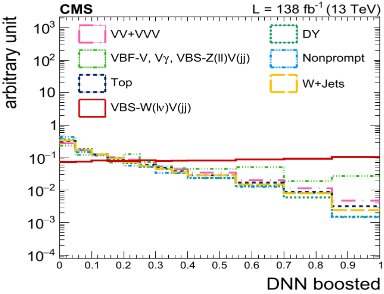
Deep Neural Network (DNN) for signal extraction

DNN training

- Dataset: 80% training, 20% validation
- Train with Adam minimizer and decaying learning rate, until full convergence
- Carefully checked and avoided overtraining







50%

Background estimation

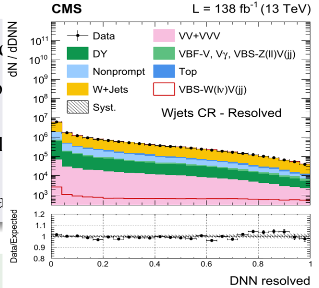
- QCD multijet background: Estimated using data-driven method
  - Tight/Loose efficiency for fake (prompt) lepton measured in QCD (Drell-Yan) enriched data sample
  - Construct relation between # prompt/ fake leptons and # passing/failing tight ID
  - Weight events in Loose data control region by (probability to have at least one non-prompt lepton) x (probability to still pass tight selection)

- W+Jets
  - Mismodelling of the jet p_T spectrum for W+many-jets sample → data-driven differential corrections
  - W+Jets contribution taken from MC but corrected in a data driven way
  - Fit their normalization in the global fit in the W+jets CR
  - After the data driven estimation → Predictions and data agree within uncertainties

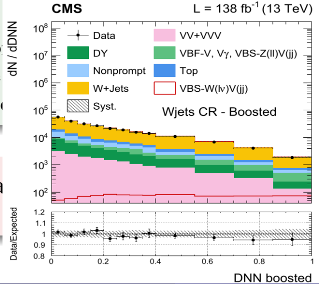
- Top background: Shape from MC, normalisation extracted from top CR in the final to fit the data

Background estimation

- QCD multijet background: Estimated using (
  - Tight/Loose ef ciency for fake (prompt) lep … riched data sample
  - Construct relation between # prompt/ fake l
  - Weight events in Loose data control region non-prompt lepton) x (probability to still pa



- W+Jets
  - Mismodelling of the jet p_T spectrum for W+many-jets sample → data-driven differential corrections
  - W+Jets contribution taken from MC but co
  - Fit their normalization in the global t in the
  - After the data driven estimation → Predicti … es

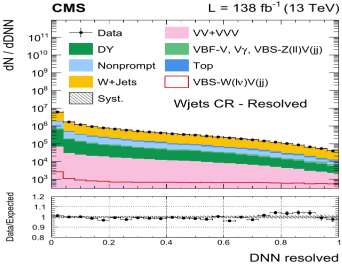


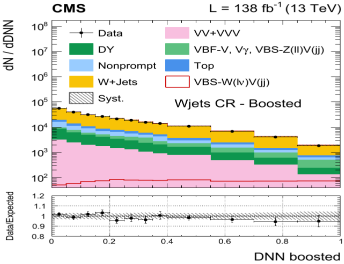
- Top background: Shape from MC, normalisa … final to fit the data

66.7%

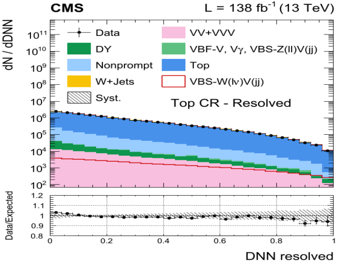
Control Regions

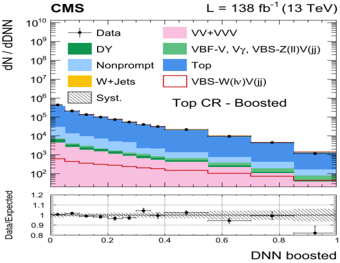
WJets CR





Top CR

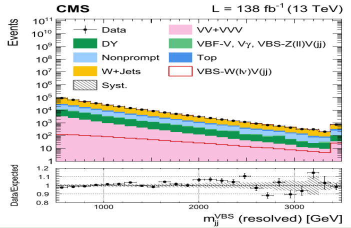


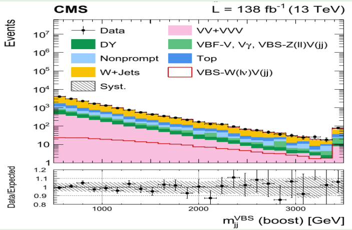


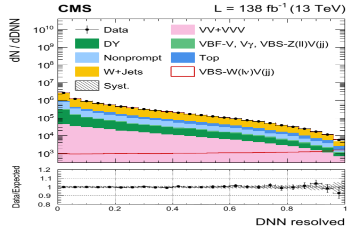
75%

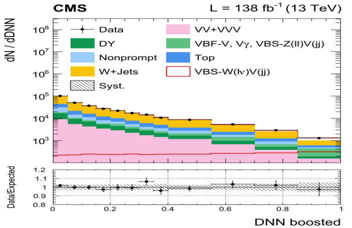
Signal Region

Signal region distribution for one of highest ranked variable, di-jet invariant mass of VBF jets and DNN score, for both resolved and boosted case









83.3%

Systematic Uncertainty

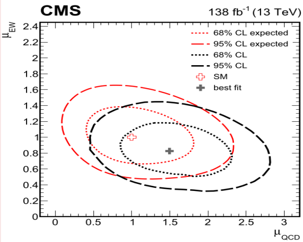
- Systematic uncertainty can affect the shape and normalisation of the DNN distribution
- Largest impact is from statistics
  - Expected as VBS signal is a rare process
- Experimental uncertainty is mainly dominated by b-tagging and jet energy scale/resolution

Uncertainty source	$\Delta\mu_{EW}$
Statistical	0.12
Limited sample size	0.10
Normalization of backgrounds	0.08
Experimental	
b-tagging	0.05
Jet energy scale and resolution	0.04
Integrated luminosity	0.01
Lepton identification	0.01
Boosted V boson identification	0.01
Total	0.06
Theory	
Signal modeling	0.09
Background modeling	0.08
Total	0.12
Total	0.22

Summary and Results

- Fit DNN shape in the signal regions
- Fit W+jets subcategories normalizations in W+jets control regions
- Fit only normalization in top-quark control regions
- Results
  - SM EW signal strength:
    $\mu_{EW} = \sigma^{obs}/\sigma^{SM} = 0.85^{+0.23}_{-0.21}$
    Signal significance of **4.4** σ (5.1 σ expected)
    - Observed cross-section of $1.9^{+0.53}_{-0.46}$ pb ($2.23^{+0.08}_{-0.11}$ pb expected)
  - Considering EW and QCD WV production as signal, the signal strength:
    $\mu_{EW+QCD} = \sigma^{obs}/\sigma^{SM} = 0.97 \pm 0.06(stat)^{+0.19}_{-0.21}(syst) = 0.97^{+0.20}_{-0.22}$
  - Measured cross-section: $16.4^{+3.5}_{-2.8}$ pb ($16.9^{+2.9}_{-2.1}$ pb expected)

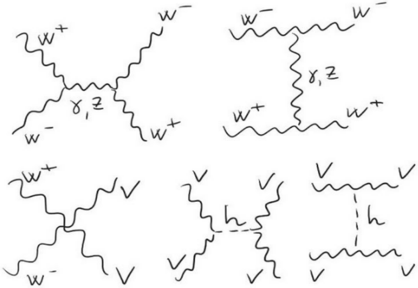
Simultaneous 2D fit of the EW and QCD WV signal strengths

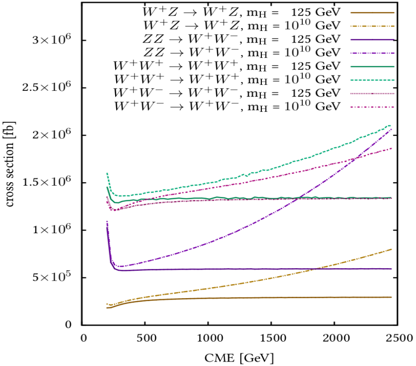


The results on the first evidence of VBS in WV channel is published in PLB

Thank you

Higgs diagrams





Back

Background Samples (1/2)

process	year	dataset name
W+Jets LO	2016	WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
	2017,2018	WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8
W+Jets LO HT bins	2016	WJetsToLNu_HT*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
	2017,2018	WJetsToLNu_HT*_TuneCP5_13TeV-madgraphMLM-pythia8
		* HT bins: 70, 100, 200, 400, 600, 800, 1200, 2500, $+\infty$ GeV
DY+Jets	2016	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
		DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
		DYJetsToLL_M-5to50_HT*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
		HT bins: 70, 100, 200, 400, 600, $+\infty$ GeV
		DYJetsToLL_M-50_HT*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
		HT bins: 70, 100, 200, 400, 600, 800, 1200, 2500, $+\infty$ GeV
	2017	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
		DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
		DYJetsToLL_M-4to50_HT*_TuneCP5_13TeV-madgraphMLM-pythia8
		HT bins: 100, 200, 400, 600, $+\infty$ GeV
		DYJetsToLL_M-50_HT*_TuneCP5_13TeV-madgraphMLM-pythia8
		HT bins: 100, 200, 400, 600, 800, 1200, 2500, $+\infty$ GeV
	2018	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
		DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
		DYJetsToLL_M-4to50_HT*_TuneCP5_13TeV-madgraphMLM-pythia8
		HT bins: 100, 200, 400, 600, $+\infty$ GeV
		DYJetsToLL_M-50_HT*_TuneCP5_13TeV-madgraphMLM-pythia8
		HT bins: 70, 100, 200, 400, 600, 800, 1200, 2500, $+\infty$ GeV
$t\bar{t}$	2016	TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
	2017, 2018	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
Semileptonic $t\bar{t}$	2016	TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
	2017, 2018	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
Single top s-channel	2016	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1
	2017, 2018	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8
Single top t-channel	2016	ST_t-channel_(anti)top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
	2017	ST_t-channel_(anti)top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8
	2018	ST_t-channel_(anti)top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
Single top tW	2016	ST_tW_(anti)top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
	2017, 2018	ST_tW_(anti)top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}$ Z production	2016	ttZJets_13TeV_madgraphMLM-pythia8
	2017, 2018	ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8
$t\bar{t}$ W production	2016	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8
	2017, 2018	ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8

Back

25%

Background Samples (2/2)

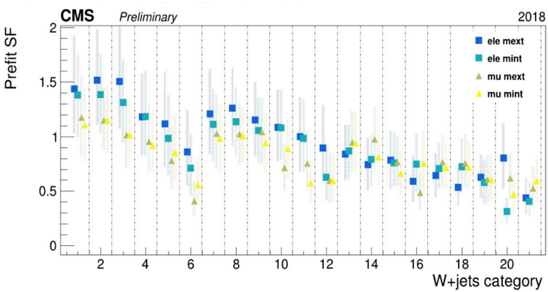
process	year	dataset name
QCD-VV	2016: TuneCUETP8M1 2017: TuneCP5 2018: TuneCP5	WminusTo2JZTo2LJJ_QCD_LO_SM_⋆ WminusToLNuWminusTo2JJJ_QCD_LO_SM_⋆ WminusToLNuZTo2JJJ_QCD_LO_SM_⋆ WplusTo2JWminusToLNuJJ_QCD_LO_SM_⋆ WplusTo2JZTo2LJJ_QCD_LO_SM_⋆ WplusToLNuWminusTo2JJJ_QCD_LO_SM_⋆ WplusToLNuWplusTo2JJJ_QCD_LO_SM_⋆ WplusToLNuZTo2JJJ_QCD_LO_SM_⋆ ZTo2LZTo2JJJ_QCD_LO_SM_⋆ ⋆ = MJJ100PTJ10_Tune⋆_13TeV-madgraph-pythia8
VVV	2016: TuneCUETP81M1 2017: TuneCP5 2018: TuneCP5	WWW_4F_Tune⋆_13TeV-amcatnlo-pythia8 WWZ_Tune⋆_13TeV-amcatnlo-pythia8 WZZ_Tune⋆_13TeV-amcatnlo-pythia8 ZZZ_Tune⋆_13TeV-amcatnlo-pythia8
Vg + VgS	2016	WGToLNuG_TuneCUETP81M1_13TeV-madgraphMLM-pythia8 ZGTo2LG_TuneCUETP81M1_13TeV-amcatnloFXFX-pythia8 WZTo3LNu_mllmin01_13TeV-powheg-pythia8
	2017, 2018	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8 ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8 WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8
VBF-V	2016	EWK_LNuJJ_MJJ-120_13TeV-madgraph-herwigpp EWK_LLJJ_MLL-50_MJJ-120_TuneEEC5_13TeV-madgraph-herwigpp_corrected
	2017, 2018	EWK_LNuJJ_MJJ-120 private production with same conf as EWK_LLJJ EWK_LLJJ_MLL-50_MJJ-120_TuneCH3_PSweights_13TeV-madgraph-herwig7_corrected
ggWW	2016, 2017, 2018	GluGluWWToLNuQQ_TuneCP5_13TeV_aMCatNLO_pythia8

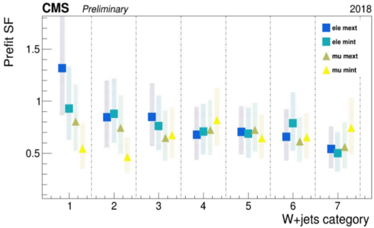
Process	Dataset name	cross section (pb)
WpTo2J_WmToLNu	/WplusTo2JWminusToLNuJJ_dipoleRecoil_EWK_LO_SM*	0.9107
WpToLNu_WmTo2J	/WplusToLNuWminusTo2JJJ_dipoleRecoil_EWK_LO_SM*	0.9114
WmToLNu_ZTo2J	/WminusToLNuZTo2JJJ_dipoleRecoil_EWK_LO_SM*	0.1000
WpToLNu_ZTo2J	/WplusToLNuZTo2JJJ_dipoleRecoil_EWK_LO_SM*	0.1825
WmTo2J_ZTo2L	/WminusTo2JZTo2LJJ_dipoleRecoil_EWK_LO_SM*	0.0298
WmToLNu_WmTo2J	/WminusToLNuWminusTo2JJJ_dipoleRecoil_EWK_LO_SM*	0.0326
WpTo2J_ZTo2L	/WplusTo2JZTo2LJJ_dipoleRecoil_EWK_LO_SM*	0.0540
WpToLNu_WpTo2J	/WplusToLNuWplusTo2JJJ_dipoleRecoil_EWK_LO_SM*	0.0879
ZTo2L_ZTo2J	/ZTo2LZTo2JJJ_dipoleRecoil_EWK_LO_SM	0.0159

2016: *MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8/RunIISummer16NanoAODv7-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/NANOAODSIM
2017: *MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v7-v1/NANOAODSIM
2018:MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8/RunIIAutumn18 NanoAODv7-Nano02Apr2020_102X_upgrade2018-_realistic_v20-v1/NANOAODSIM

37.5 %

Correction factors

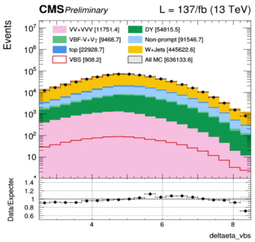


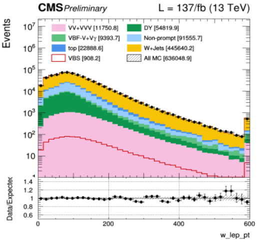


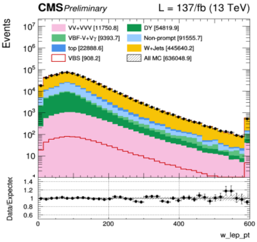
Normalization correction factors in the external and internal control regions are compatible

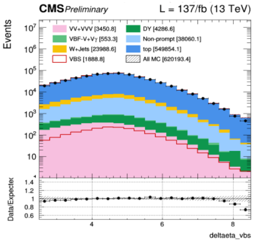
Back

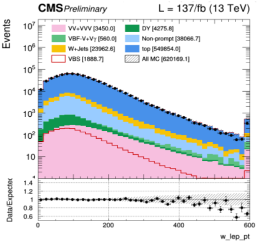
Post Fit distributions

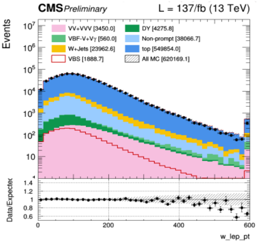






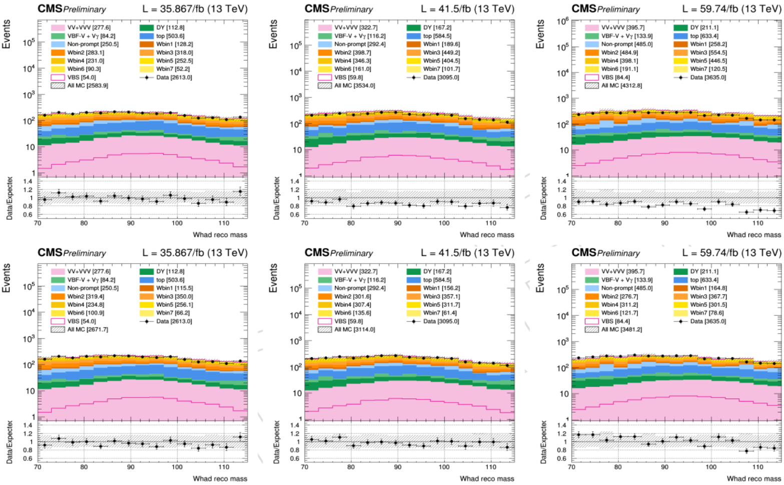




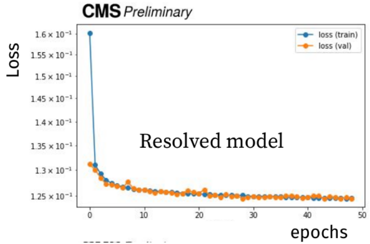


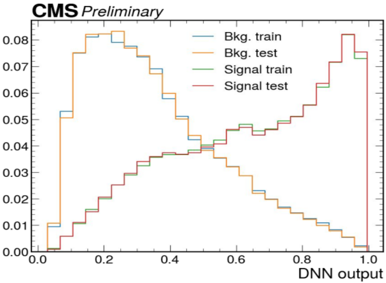
62.5%

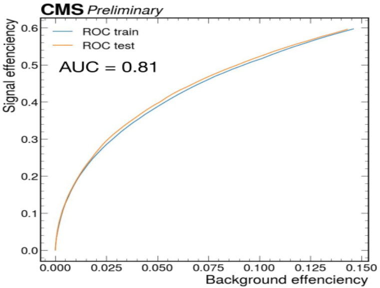
W had reco signal muon



Deep Neural Network (DNN) for signal extraction







Back

87.5%

DNN input variables

Variable	Resolved	Boosted
Lepton pseudorapidity	✓	✓
Lepton transverse momentum	✓	✓
Zeppenfeld variable for the lepton	✓	✓
Number of jets with $p_T > 30$ GeV	✓	✓
VBS leading tag-jet p_T	-	✓
VBS trailing tag-jet p_T	✓	✓
Pseudorapidity interval between VBS tag-jets	✓	✓
Quark Gluon discriminator of the highest p_T jet of the VBS tag-jets	✓	✓
Azimuthal angle distance between VBS tag-jets	✓	✓
Invariant mass of the VBS tag-jets pair	✓	✓
p_T of jets from V_{had}	✓	-
Pseudorapidity difference between V_{had} jets	✓	-
Quark Gluon discriminator of the V_{had} jets	✓	-
V_{had} p_T	-	✓
Invariant mass of the V_{had}	✓	✓
Zeppenfeld variable for the V_{had}	-	✓
V_{had} centrality	-	✓