Search for vector boson scattering production of VW pair in the semi-leptonic channel at CMS

Flavia Cetorelli

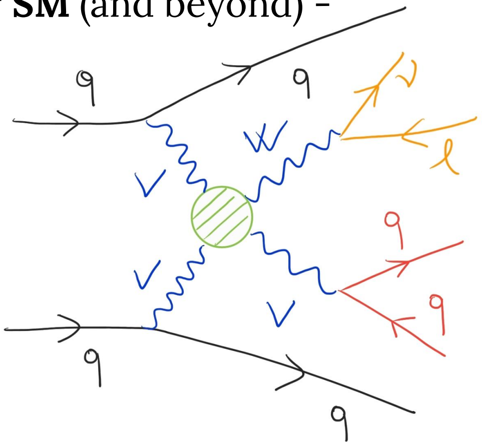
INFN & Università degli Studi di Milano - Bicocca

14/07/2021

Introduction and motivation

★ **VBS VV semileptonic channel** not yet observed at LHC.

- Powerful tools to **investigate the electroweak sector of SM** (and beyond) - connection to **Higgs boson**



- It has a **good balance** between:
  - larger **cross-section** than fully-leptonic one
  - larger **background** from V+jets processes

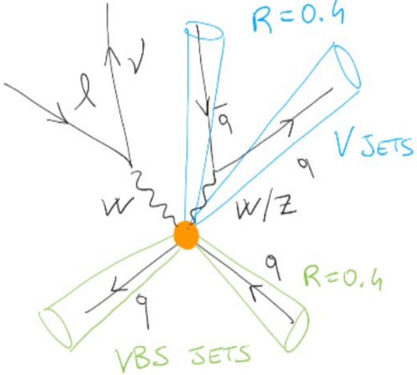
- **Multivariate techniques** and **data-driven strategies** implemented to reduce and control the overwhelming background.

Final state topology

VBS processes are mediated by **electroweak** interaction, hence in final state:

★ two jets (called **tag jets**) from initial state partons
  - with **high invariant mass**
  - **great pseudorapidity separation**

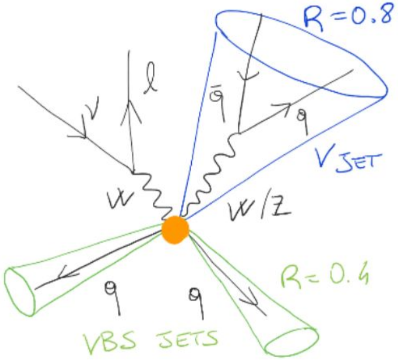
★ two **vector bosons (WV)**
  - central wrt **tag jets**;
  - **W** → leptonic decay in a electron or a muon (+ neutrino)
  - **V** → hadronic decay in:



Two jets: detector able to **resolve** the two as separated objects

OR

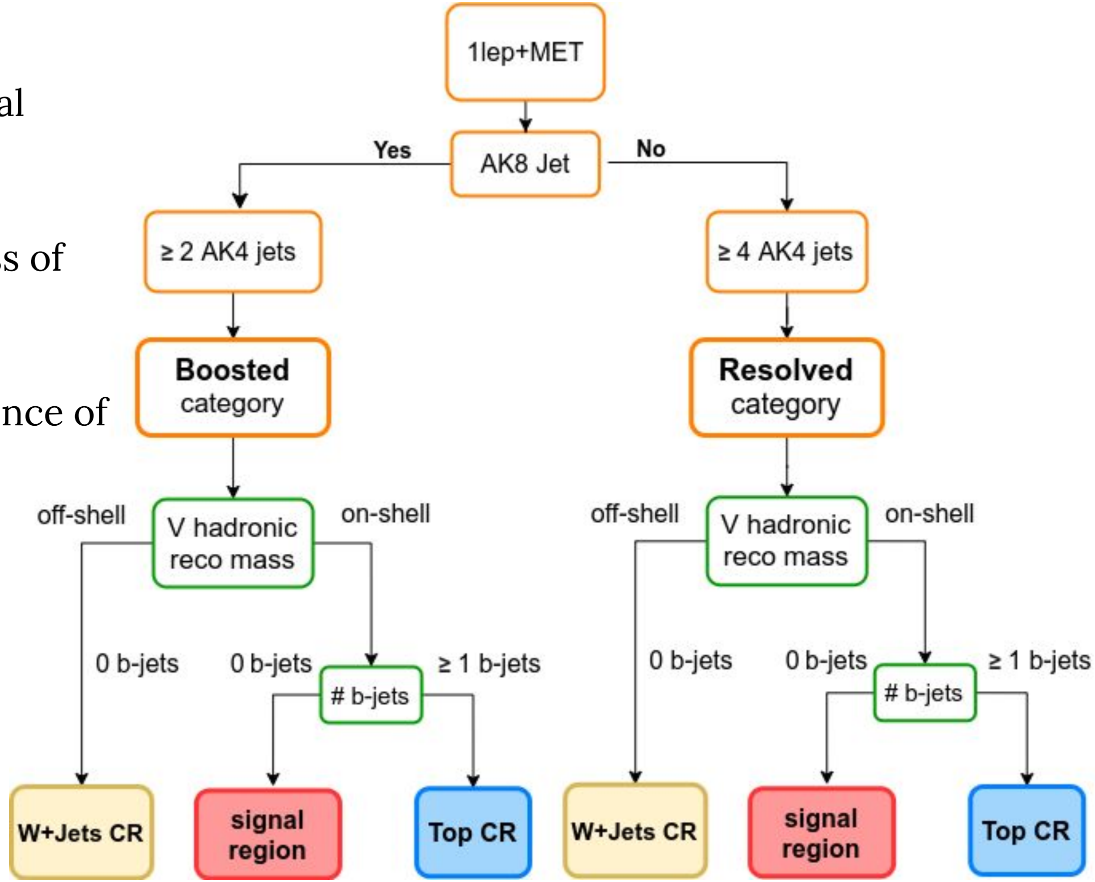
One large-radius jet: when the V is **boosted** the two jets are reconstructed as **one** large-radius jet, with **high pT** (> 200 GeV)

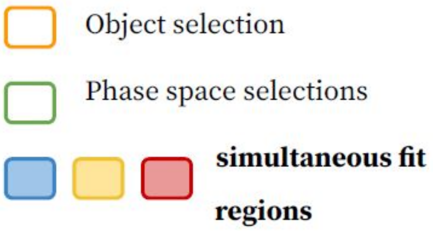


Analysis phase space

Control Regions are **orthogonal** to Signal Region:

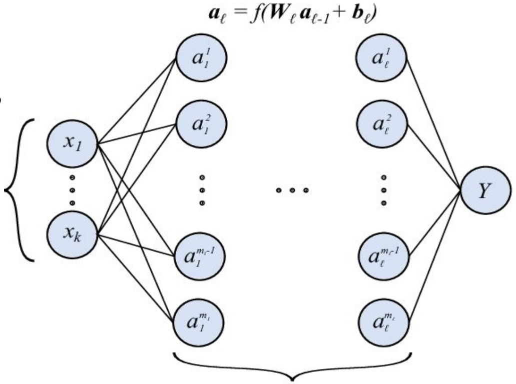
- ★ W+Jets **CR**: defined wrt to the mass of the hadronically decaying V
- ★ **Top-quark CR:** requiring the presence of b jets from top decay.





Deep Neural Network

- ★ **Electroweak VBS** signal **vs** all the **backgrounds**
- ★ Feed-forward deep neural network (**DNN**) as discriminator:
  - **Two** different models
    - **Boosted** phase space: DNN fully connected (64-32-32-32 nodes) + 13 inputs
    - **Resolved** phase space: DNN fully connected (64-64-64-64 nodes) + 16 inputs
  - Joining **all years** together
  - Samples split in 80% for training - 20% for validation
- ★ Overfitting is carefully avoided thanks to regularization techniques



Backgrounds estimation

- **Main backgrounds:**
  - Top : MC simulation for shape + dedicated **control region** to fit the normalization.
  - W+jets : MC simulations + data-driven corrections:
    - Taking care of the **poor MC description**, in particular of p_T of W leptonically decaying and p_T of one tag jet
    - W+jets split into several components according to the 2 variables: normalization unconstrained and correlated during the fit (**measured in the CR**!)
- **Minor backgrounds:**
  - **QCD-VV**, **VBF-V**, **DY, ggWW, VVV, Vγ:** from **MC simulation**
  - **Non-prompt**: data-driven estimation with fakable object technique

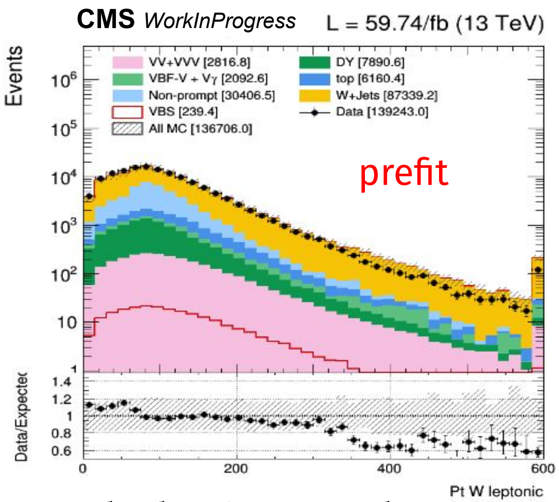
W+jets background estimation

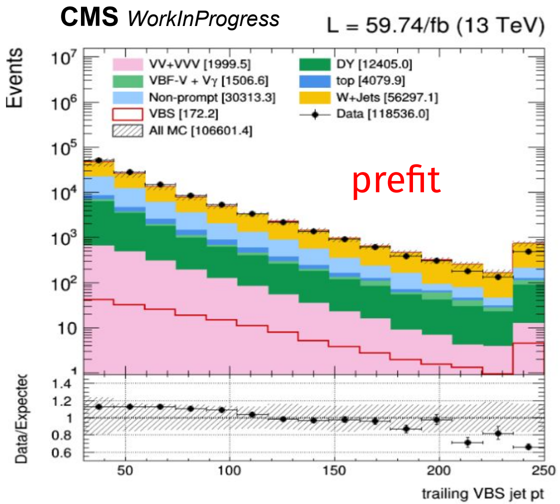
Differential correction to W+jets:

1. Split W+jets background MC sample in bins of:
   a. **leptonic W p_T** and **trailing VBS jet p_T** → resolved category only
   b. **leptonic W p_T** → boosted category only
2. **Fit their normalization in the global fit** in the W+jets control regions

1D binning in **boosted** category

Bin	W_{lep}^{pt} bin
1	$W_{lep}^{pt} < 50$ GeV
2	$50 \leq W_{lep}^{pt} < 100$ GeV
3	$100 \leq W_{lep}^{pt} < 150$ GeV
4	$150 \leq W_{lep}^{pt} < 200$ GeV
5	$200 \leq W_{lep}^{pt} < 300$ GeV
6	$300 \leq W_{lep}^{pt} < 400$ GeV
7	$W_{lep}^{pt} \geq 400$ GeV





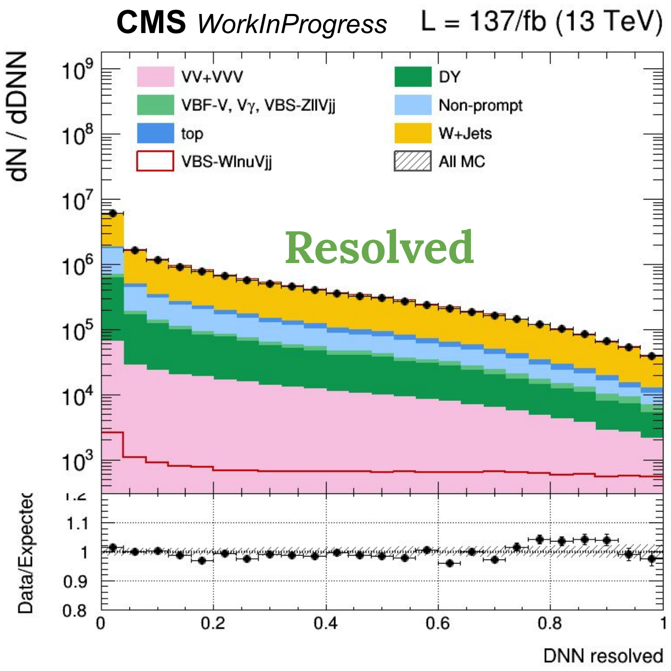
Resolved W+jets control region (2018, electron channel)

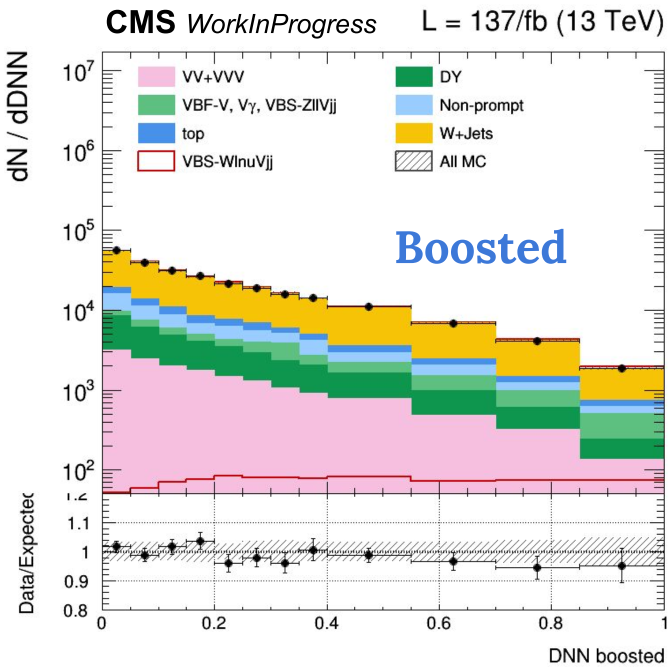
2D binning in **resolved** category

Bin	W_{lep}^{pt} bin	$VBS^2_{p_T}$ bin
1		$VBS^2_{p_T} < 55$ GeV
2		$55 \leq VBS^2_{p_T} < 75$ GeV
3		$75 \leq VBS^2_{p_T} < 100$ GeV
4	$W_{lep}^{pt} < 100$ GeV	$100 \leq VBS^2_{p_T} < 135$ GeV
5		$135 \leq VBS^2_{p_T} < 170$ GeV
6		$VBS^2_{p_T} >= 170$ GeV
7		$VBS^2_{p_T} < 55$ GeV
8		$55 \leq VBS^2_{p_T} < 75$ GeV
9		$75 \leq VBS^2_{p_T} < 100$ GeV
10	$100 \leq W_{lep}^{pt} < 200$ GeV	$100 \leq VBS^2_{p_T} < 135$ GeV
11		$135 \leq VBS^2_{p_T} < 170$ GeV
12		$VBS^2_{p_T} >= 170$ GeV
13		$VBS^2_{p_T} < 90$ GeV
14		$90 \leq VBS^2_{p_T} < 125$ GeV
15	$200 \leq W_{lep}^{pt} < 300$ GeV	$125 \leq VBS^2_{p_T} < 160$ GeV
16		$VBS^2_{p_T} >= 160$ GeV
17	$300 \leq W_{lep}^{pt} < 400$ GeV	$VBS^2_{p_T} < 90$ GeV
18		$VBS^2_{p_T} >= 90$ GeV
19	$400 \leq W_{lep}^{pt} < 500$ GeV	$VBS^2_{p_T} < 85$ GeV
20		$VBS^2_{p_T} >= 85$ GeV
21	$W_{lep}^{pt} \geq 500$ GeV	

W+jets control regions

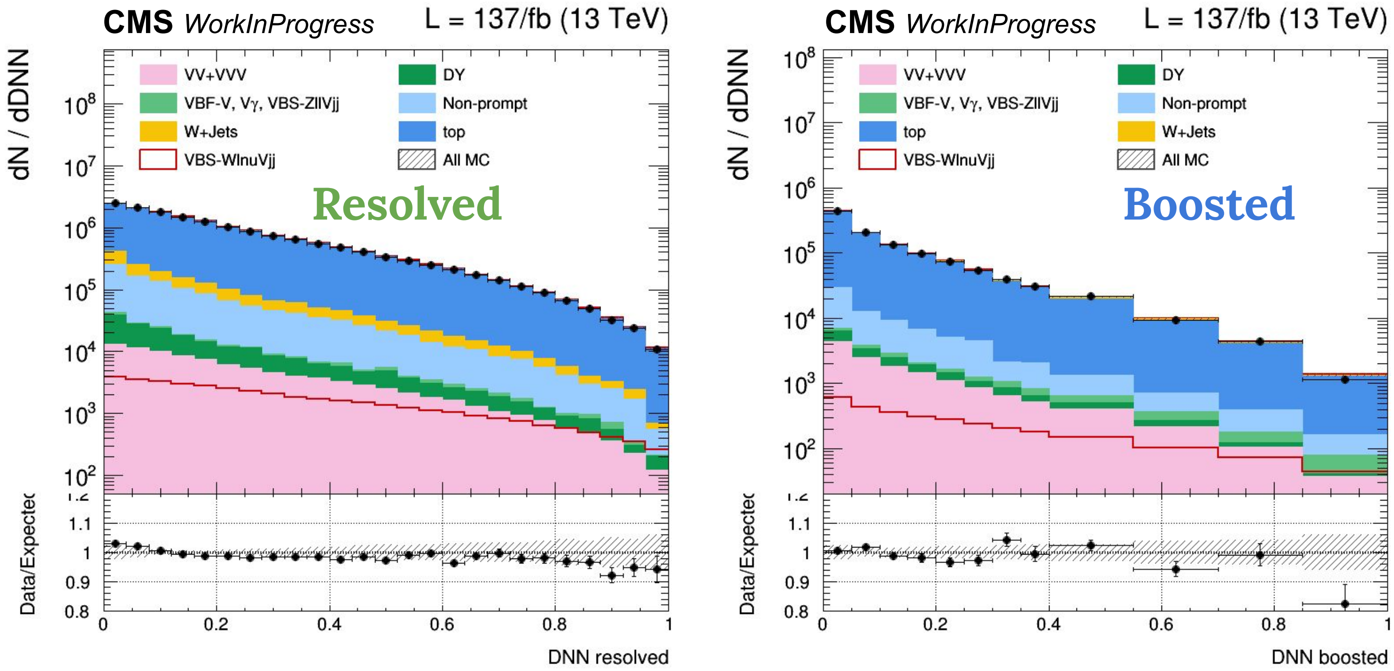
After the data driven estimation → Good agreement in both categories.





Top-quark control regions

After the data driven estimation → Good agreement in both categories.

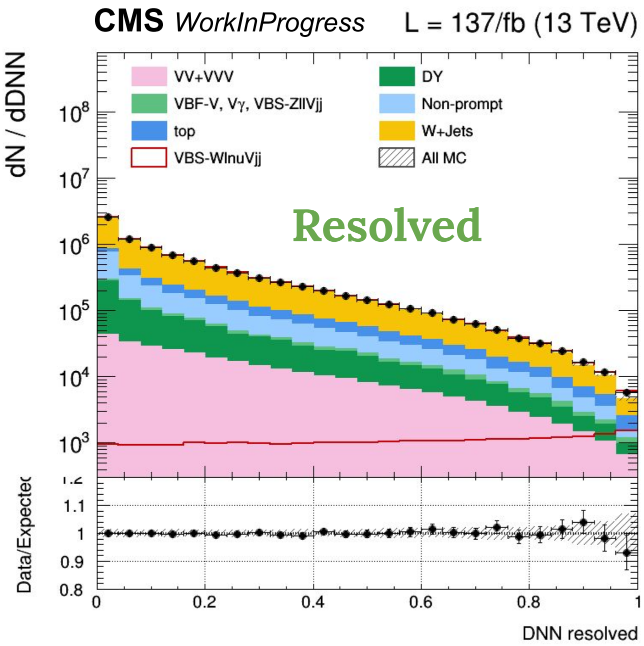


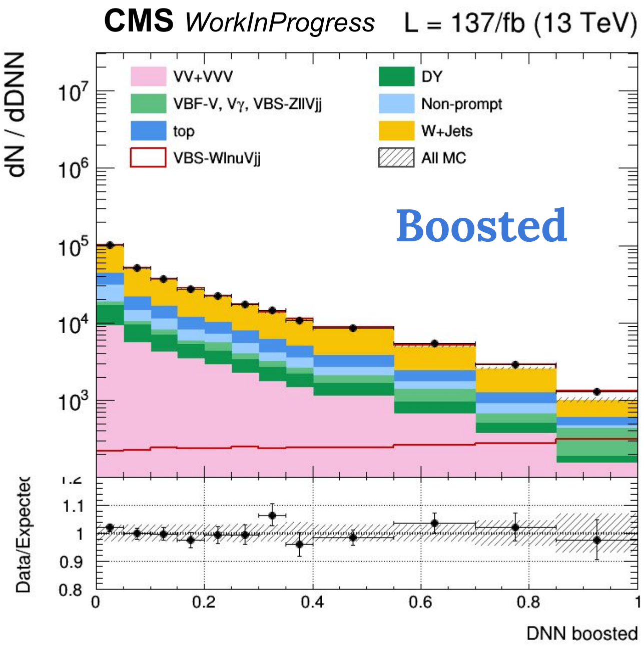
Signal extraction

To extract the final signal strength:

- ★ **the DNN**s are fitted in the signal phase spaces,
- ★ **the yield** in the control regions → normalize the W+jets and top backgrounds.

Post-fit plot for signal regions, after data driven corrections





Results

- **EW WV** measurements:
  - The **observed (expected) significance** is **4.4 (5.1) σ**
  - **Signal strength modifier** is: $\mu_{EW} = \sigma_{obs}/\sigma_{SM} = 0.85^{+0.24}_{-0.20}$
  - **Cross section** in the fiducial phase space: $1.91\ \text{pb}^{+0.53}_{-0.45}$

    expected $2.23\ \text{pb}^{+0.08}_{-0.11}$(scales) $^{+0.05}_{-0.05}$(pdf)

Statistically limited analysis
Run III is going to **improve** the results

- **EW + QCD WV** production:
  - **Signal strength modifier** is: $\mu_{EW+QCD} = 0.98^{+0.20}_{-0.17}$
  - **Cross section** in the fiducial phase space: $16.6\ \text{pb}^{+3.4}_{-2.9}$

    expected $16.9\ \text{pb}^{+2.9}_{-2.1}$(scale) $^{+0.5}_{-0.5}$(pdf)

Fiducial phase space definition
Parton level cuts:
- all partons $p_T > 10$ GeV
- at least one pair with $m_{qq} > 100$ GeV

Summary

- ★ Search for **VBS WV semileptonic channel** with the CMS detector at LHC
  - Reveal more insight on the **electroweak sector of SM**
  - **Multivariate techniques** and **data-driven strategies** implemented to reduce and control the overwhelming background
  - **First evidence** of VBS WV semileptonic at LHC:
    - observed (expected) significance: 4.4 (5.1) σ
    - signal strength $\mu_{EW} = \sigma_{obs}/\sigma_{SM} = 0.85^{+0.24}_{-0.20}$
  - Performed also fit to extract **EW+QCD**:
    - $\mu_{EW+QCD} = 0.98^{+0.20}_{-0.17}$

Backup

Backgrounds

- **Main backgrounds:**
  - **W+jets**: Madgraph+Pythia8, HT binned LO samples → data-driven correction
  - **Top**: Powheg+Pythia8, ttbar, single top, tW, tZ

- **Minor backgrounds:**
  - **QCD-VV**: Madgraph+Pythia8, same parton level cuts as signal
  - **Non-prompt**: data-driven estimation with fakable object technique
  - **VBF-V**: single V boson EWK production, Madgraph+Herwig7
  - **DY:** aMC@NLO + HT binned LO
  - **ggWW, VVV, Vγ:** very small contribution

Selections

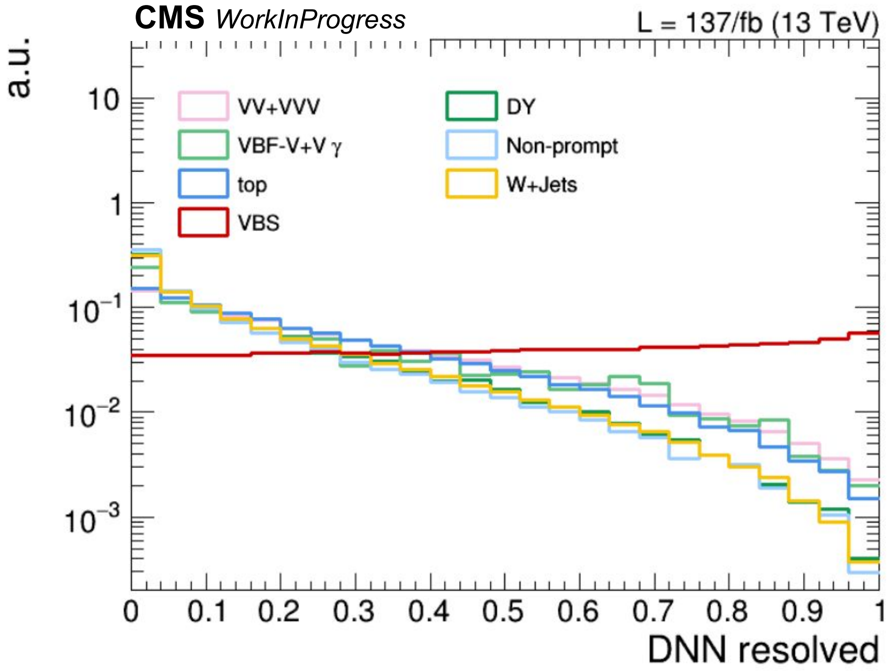
	Signal region	Top control region	W+jets control region
Resolved category	Ele $p_T >$ 30 GeV (2016), 35 GeV (2017, 2018) Muon $p_T >$ 30 GeV PuppiMET > 30 GeV Leading VBS jet $p_T >$ 50 GeV trailing VBS jet and Vjets $p_T >$ 30 GeV $\Delta\eta_{VBS} > 2.5$, $M_{jj\ VBS} > 500$ GeV Leptonic $M^T_W < 185$ 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**

Input variables DNN

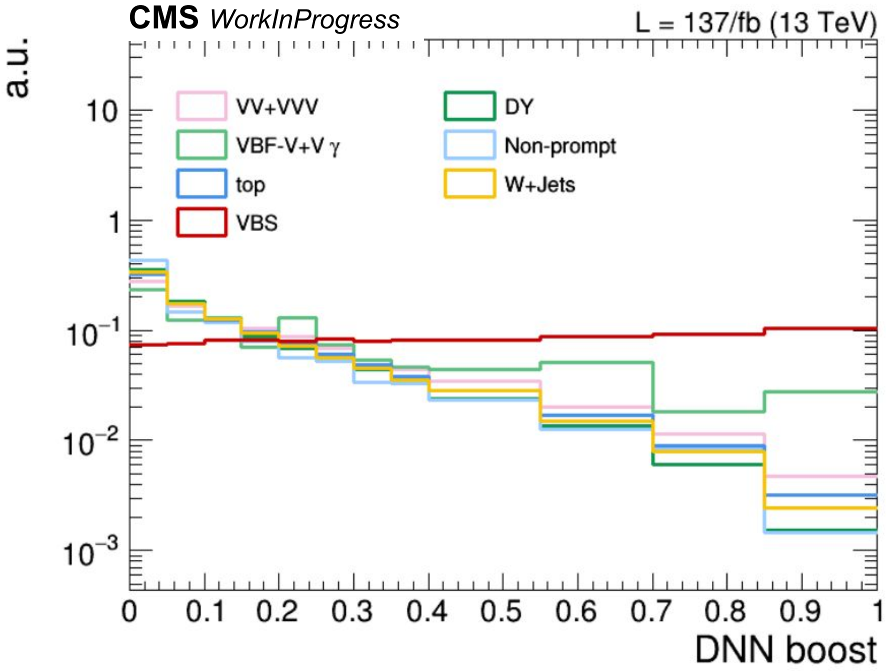
Variable	Resolved	Boosted
Lepton pseudorapidity	x	x
Lepton transverse momentum	x	x
Zeppenfeld variable for the lepton	x	x
Number of jets with $p_T > 30$ GeV	x	x
VBS tag-jets p_T	x	x
Pseudorapidity interval between VBS tag-jets	x	x
Quark Gluon discriminator of the highest p_T jet of the VBS tag-jets	x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

Discriminator output

Resolved



Boosted



Uncertainties breakdown

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