

Search for vector boson scattering with the semi-leptonic WV signature at CMS

Davide Valsecchi on behalf of the CMS Collaboration

Università degli Studi di Milano-Bicocca, CERN

E-mail: davide.valsecchi@cern.ch

A search for the electroweak VBS production of a WV pair plus two jets, in the semi-leptonic channel, at a center-of-mass energy of 13 TeV is reported. The data sample corresponds to the full Run-II CMS dataset of proton-proton collisions at 13 TeV including an integrated luminosity of 137.1 fb^{-1} . Events are analyzed in two energy regimes: either the hadronically decaying W/Z boson is reconstructed as one large-radius jet, or it is identified as a pair of jets with dijet mass near the W/Z mass. Machine learning models are optimized for the signal extraction and the classifiers are interpreted using tools from the explainable machine learning field. The overwhelming background contribution from the single W production plus jets is measured in dedicated control regions implementing a data-driven strategy.

The European Physical Society Conference on High Energy Physics (EPS-HEP2021),

26-30 July 2021,

Online conference, jointly organized by Universität Hamburg and the research center DESY

1. Introduction

The observation of a Higgs boson with a mass of about 125 GeV by the CMS and ATLAS collaborations [1, 2] consolidated the validity of the Standard Model (SM) of elementary interactions: within the current uncertainties, the newly found particle is consistent with the SM expectations. Nevertheless, several fundamental issues remain unsolved. In particular, new phenomena may be present in the electroweak sector, where a sensitive probe of new physics is naturally given by the study of the scattering at high energy of massive electroweak bosons, usually known as Vector Boson Scattering (VBS) [3]. VBS happens at the Large Hadron Collider (LHC) when a quark from each of the incoming protons radiates a vector boson W or Z. These two vector bosons interact with each other, and their decay products are detected in the LHC experiments. At leading-order (LO) in perturbation theory VBS is a purely $2 \rightarrow 6$ fermions electroweak processes of order $\mathcal{O}(\alpha_{\text{EW}}^6)$.

The target of this analysis is the extraction of the WV VBS production cross section, signal strength $\mu = \sigma^{\text{obs}}/\sigma^{\text{SM}}$ (where σ^{obs} and σ^{SM} are the observed and predicted cross section respectively), and its corresponding significance, in the semileptonic WV channel, where one final state vector boson V decays hadronically and the other (W) leptonically. The search is performed on the data set of proton-proton collisions at 13 TeV collected with the CMS detector [4] during 2016-2018, corresponding to an integrated luminosity of 137 fb^{-1} . The largest background contribution is given by the single W boson production process in association with jets, followed by the top quark and the QCD multijet backgrounds. The QCD WV production irreducible background, of order $\mathcal{O}(\alpha_{\text{EW}}^4 \alpha_S^2)$, is separable from VBS only at LO in the perturbation expansion [5] and it is reduced considerably in the analysis by the phase space preselections. A simultaneous measurement of the EW and QCD WV production signal strengths is also performed.

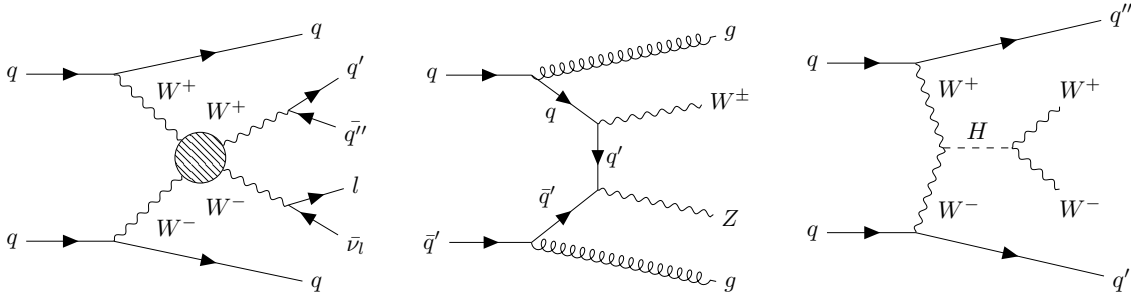


Figure 1: Examples of Feynman diagrams contributing to the analyzed final state: purely EW VBS process contributions (left diagram), QCD diboson production (center diagram), and s-channel Higgs boson contribution (right diagram).

The signal, namely the VBS $W(l\nu)V(jj)$ process, is simulated with the MADGRAPH5_AMC@NLO v2.6.5 Monte Carlo (MC) at LO precision and decayed with MadSpin [6]. The backgrounds are simulated with POWHEG v2 [7], MADGRAPH5_AMC@NLO v2.4.2 [8], and parton shower, hadronization, and the simulation of the underlying event are provided by PYTHIA 8.226 (8.230) [9]. The dipole recoil scheme is used in the parton shower PYTHIA simulation for the VBS signal MC sample, in order to describe correctly the additional jet emissions in the VBS topology [5, 10].

2. Event selection

Events are selected requiring the presence of a single isolated electron (muon) with $p_T > 35$ GeV (30 GeV) and missing transverse momentum $p_T^{\text{miss}} > 30$ GeV, and either four small radius (named AK4) hadronic jets or 2 AK4 and 1 large radius jet (named AK8), clustered using the anti- k_T algorithm with a distance parameter of 0.4 for AK4 jets or 0.8 for AK8 ones. One pair of AK4 jets with $p_T > 50(30)$ GeV, called VBS tag jets, is used to tag the event requiring a large invariant mass $m_{jj}^{\text{VBS}} > 500$ GeV and large pseudorapidity separation $\Delta\eta_{jj}^{\text{VBS}} > 2.5$, the typical signature of VBS-like events. Between the remaining jets, the decay products of the hadronically-decaying final state vector boson W or Z, must be identified to fully reconstruct the VBS topology. The signal efficiency is maximized targeting different energetic regimes: If the V boson is boosted ($p_T > 200$ GeV) its decay products are reconstructed in the detector as a single energetic large radius AK8 jet. Experimentally, the boosted V boson decay is identified using the ratio between 2- and 1-subjettiness [11] variables denoted as $\tau_{21} = \tau_2/\tau_1 < 0.45$ and groomed AK8 jet mass between 40 and 250 GeV [12]. At smaller p_T , the V boson is reconstructed as a pair of smaller radius jets with invariant mass near the W or Z boson mass.

The contamination from the top quark backgrounds is reduced drastically in the signal region by vetoing b-jets using the CMS multivariate DeepCSV [13] tagger. Inverting the b-veto requirement a control region for the top quark backgrounds is defined and included in the fit. The W+jets background is reduced selecting as signal region a narrow invariant mass interval for the hadronic vector boson decays products around the W or Z mass peak ($m_V \in (65, 105)$ GeV). The invariant mass sidebands are used to define a control region for the W+jets background used in the fit.

3. Signal discrimination

Because of the large background on the one hand, and of the sophisticated signal topology on the other, the most significant features that allow to separate them have been combined in two discriminators built with deep neural networks (DNN), one for the resolved and one for the boosted categories. The DNN is implemented as a fully connected neural network with 4 layers with 64 (32) nodes for the resolved (boosted) topology, trained via stochastic gradient descent. The most discriminating input variables are m_{jj}^{VBS} and the Zeppenfeld variable of the lepton [14]. A technique called SHAP (SHapley Additive exPlanations) [15], developed in the field of explainable machine learning, is applied to cross check the dependence of the DNN model on the input variables and to rank their importance.

4. Results

The signal extraction is performed using the modified frequentist likelihood ratio profiling approach. Experimental and theoretical uncertainties are represented by nuisance parameter that morphs the shapes of the distributions for the signal and background processes or scales their total normalization. The top quark background normalization is left free to float in the fit. The W+jets background is also corrected in a data-driven way: its MC sample is split in coarse set of bins, and the normalization of each sub-category is left free to float in the fit allowing to correct differentially for residual discrepancies in its shape between data and Monte Carlo.

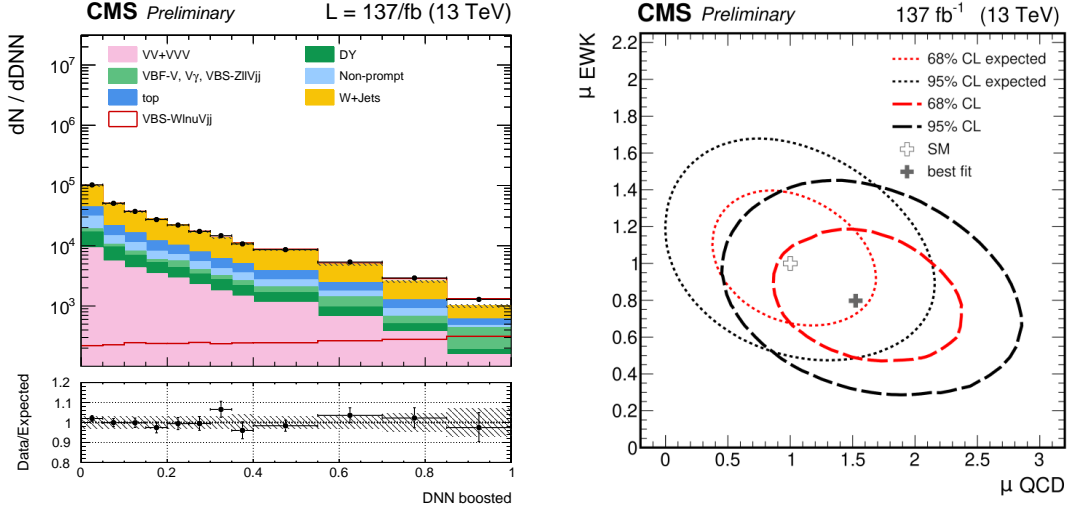


Figure 2: Left: post-fit DNN distributions for the boosted signal region, for the EW-signal-only fit. Right: 68% and 95% confidence level contours for the simultaneous fit of EW and QCD WV signal strengths. [16]

Three separate interpretations are performed on the selected events: the measurement of the purely EW signal strength μ_{EW} keeping the QCD WV production contribution fixed to the SM prediction $\mu_{QCD} = 1$; the measurement of the signal strength of the sum of the EW and QCD WV processes; a two-dimensional simultaneous measurement of the signal strengths μ_{EW} and μ_{QCD} .

A fiducial phase space region is defined at parton level requiring all partons to have $p_T > 10$ GeV and at least one pair of outgoing partons with invariant mass $m_{qq} > 100$ GeV. The measured (SM expected) EW WV production cross section is $1.90_{-0.47}^{+0.54}$ pb ($2.23_{-0.11}^{+0.08}$ (scale) ± 0.05 (PDF) pb), corresponding to an observed EW-only signal strength of: $\mu_{EW} = 0.85_{-0.20}^{+0.24} = 0.85 \pm 0.12$ (stat) $_{-0.17}^{+0.21}$ (syst), $1_{-0.22}^{+0.24}$ expected. The observed (expected) significance for the SM EW WV signal is 4.4 (5.1) standard deviations.

Considering instead the signal as the sum of the EW and QCD WV production processes, the cross-section is measured (expected) to be $16.6_{-2.9}^{+3.4}$ pb ($16.9_{-2.1}^{+2.9}$ (scale) ± 0.5 (PDF) pb), in the same fiducial phase space region as the EW-only one. The total signal strength, $\mu_{EW+QCD} = 1_{-0.18}^{+0.19}$ expected, is measured as: $\mu_{EW+QCD} = 0.98_{-0.17}^{+0.20} = 0.98 \pm 0.07$ (stat) $_{-0.16}^{+0.19}$ (syst). The fit is performed also leaving as free independent parameters the signal strength of the EW and QCD WV production components (μ_{EW} and μ_{QCD}), as shown in Fig. 2.

5. Conclusion

Thanks to the large Run II dataset delivered by LHC and recorded by CMS, sophisticated multivariate techniques to extract the signal, and careful data-driven corrections of the overwhelming backgrounds, the first evidence for semileptonic WV VBS at LHC has been reached and presented. The measured signal strengths are in agreement with the SM predictions within the 68% confidence interval.

References

- [1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [1207.7214].
- [2] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235].
- [3] C.F. Anders et al., *Vector boson scattering: Recent experimental and theory developments*, *Rev. Phys.* **3** (2018) 44 [1801.04203].
- [4] CMS collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [5] R. Covarelli, M. Pellen and M. Zaro, *Vector-boson scattering at the LHC: unraveling the electroweak sector*, *Int. J. Mod. Phys. A* **36** (2021) 2130009 [2102.10991].
- [6] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015 [1212.3460].
- [7] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581].
- [8] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].
- [9] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].
- [10] A. Ballestrero et al., *Precise predictions for same-sign W -boson scattering at the LHC*, *Eur. Phys. J. C* **78** (2018) 671 [1803.07943].
- [11] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N -subjettiness*, *JHEP* **03** (2011) 015 [1011.2268].
- [12] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [1402.2657].
- [13] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].
- [14] D.L. Rainwater, R. Szalapski and D. Zeppenfeld, *Probing color singlet exchange in $Z + two jet$ events at the CERN LHC*, *Phys. Rev. D* **54** (1996) 6680 [hep-ph/9605444].
- [15] S.M. Lundberg and S.-I. Lee, *A unified approach to interpreting model predictions*, NIPS'17, (Red Hook, NY, USA), p. 4768, Curran Associates Inc., 2017 [1705.07874].
- [16] CMS collaboration, *Search for vector boson scattering at the LHC Run 2 with CMS data in the semi-leptonic $lvqq$ final state*, Tech. Rep. CMS-PAS-SMP-20-013, CERN, Geneva (2021).