

Differential measurement of the Higgs boson production with two jets in the WW decay channel

B. CAMAIANI

ON BEHALF OF THE CMS COLLABORATION

*Department of Physics and Astronomy, Via G. Sansone 1, 50019 Sesto Fiorentino (FI), Italy
INFN Firenze, Via G. Sansone 1, 50019 Sesto Fiorentino (FI), Italy*

Summary. — A measurement of the differential production cross section of the Higgs boson decaying into two W bosons, with two jets produced in association, is presented. The analysis is based on proton-proton collision data collected by the CMS detector from 2016 to 2018. A novel approach based on adversarial neural networks is employed to reduce model dependence and enable reinterpretability. Constraints on several Wilson coefficients are also derived.

1. – Introduction

The discovery of the Higgs boson (H) by the ATLAS and CMS [1] Collaborations has initiated a series of precision measurements with the aim of determining the properties of the newly observed particle. So far, no significant deviations from the predictions of the Standard Model (SM) have been observed. While current data show that Higgs boson couplings to gauge bosons align with the SM, deviations due to anomalous couplings (AC) are still possible. In beyond-the-SM (BSM) scenarios, such AC contributions could alter the HVV interaction structure (where V is a vector boson), introducing CP-even or CP-odd effects that may lead to observable changes in the kinematics of final-state particles. A simultaneous presence of CP-odd AC and SM terms could also induce CP violation in Higgs boson interactions.

One of the most sensitive observables for detecting BSM effects in the HVV interaction is the signed azimuthal angle difference ($\Delta\Phi_{jj}$) between the two leading jets produced in association with the Higgs boson in proton-proton (pp) collisions. These jets can originate either from additional QCD radiation in gluon gluon fusion (ggF) production or from the vector boson fusion (VBF) process. Following the definition provided in Ref. [2], the $\Delta\Phi_{jj}$ observable is given by

$$(1) \quad \Delta\Phi_{jj} = \phi_{jk} - \phi_{jl} \quad \text{with} \quad \eta_{jk} > \eta_{jl} ,$$

where η_{jk} and η_{j1} are the pseudorapidities of the two jets with the highest transverse momenta, j_k and j_1 , in the event.

This note presents a differential measurement of the Higgs boson production cross section as a function of $\Delta\Phi_{jj}$ in the $H \rightarrow WW$ decay mode, using pp collision data collected by the CMS detector from 2016 to 2018 at $\sqrt{s} = 13$ TeV, corresponding to a total integrated luminosity of 138 fb^{-1} [3]. The analysis selects events with a different-flavor dilepton final state from $H \rightarrow WW \rightarrow 2\ell 2\nu$ decays and is carried out within a fiducial volume to avoid extrapolation to the full phase space based on underlying theoretical assumptions. An unfolding procedure is employed to correct for detector effects, ensuring that the result can be directly compared to any physics model.

Although the unfolded results are model-independent, model assumptions enter through the signal extraction procedure, which typically relies on Monte Carlo (MC) simulations of the SM signal process as a benchmark. This can bias the reinterpretation of results under BSM scenarios, since the shape of the fit variable distribution generally depends on Higgs boson properties, such as the HVV couplings. To mitigate this, a domain adaptation (DA) technique is employed to construct a model-agnostic fit variable, enhancing the robustness of the measurement. The resulting cross section is then reinterpreted within the SM effective field theory (SMEFT) framework to constrain Wilson coefficients in the Warsaw basis [4].

2. – Event categorization

Events are selected by requiring that the two highest- p_T leptons (referred to as the leading and subleading lepton candidates) have opposite electric charge and different flavor ($e\mu$), in order to suppress the Drell–Yan (DY) background. The leading lepton is required to have $p_T > 25$ GeV, while the subleading lepton must have $p_T > 13$ GeV. In the 2016 data set, this threshold is lowered to 10 GeV for muons, due to differences in the high-level trigger (HLT) requirements compared to those used in 2017 and 2018. To reduce contributions from minor backgrounds such as WZ and triboson production, any event with a third lepton with $p_T > 10$ GeV is vetoed. The invariant mass ($m_{\ell\ell}$) and the transverse momentum ($p_T^{\ell\ell}$) of the dilepton system must be greater than 12 GeV and 30 GeV, respectively. Owing to the presence of neutrinos in the final state, a minimum threshold of 20 GeV is imposed on the missing transverse momentum (E_T^{miss}). The 2-jet final state is then selected by requiring at least two jets, each with $p_T > 30$ GeV, and an invariant mass of the dijet system (m_{jj}) greater than 120 GeV.

After the preselection, the dominant background contributions arise from $t\bar{t} + tW$ and $DY \rightarrow \tau\tau$ processes. To constrain the normalization of these backgrounds, two dedicated control regions (CRs) are defined. The top CR is designed to be enriched in top quark events, while the DY CR targets events from $DY \rightarrow \tau\tau$ production. Finally, a signal region (SR) is defined to enhance sensitivity to signal events. The SR is constructed by requiring $m_T > 60$ GeV and $m_T^{\ell_2} > 30$ GeV, where m_T and $m_T^{\ell_2}$ are transverse mass variables built from E_T^{miss} and the dilepton or subleading lepton momenta, respectively. Events containing b-tagged jets, identified using a loose b-tag requirement, are vetoed to suppress top-quark backgrounds. The top CR is defined by requiring at least one b-tagged jet with $p_T > 20$ GeV and $m_T^{\ell_2} > 30$ GeV, with an increased $m_{\ell\ell}$ threshold of 50 GeV. The DY CR restores the b-jet veto and inverts the m_T requirement, additionally requiring $40 < m_{jj} < 80$ GeV, with no condition on $m_T^{\ell_2}$. All CRs and the SR are divided into four equally-spaced bins in the $\Delta\Phi_{jj}$ observable, covering the range from $-\pi$ to π .

The Higgs boson differential cross section is measured within a fiducial phase space,

defined at generator level to minimize the dependence of the measurement on the assumed Higgs boson production model. At this level, the same bin boundaries used at reconstruction level are applied to the $\Delta\Phi_{jj}$ distributions.

3. – Model-agnostic classification

To enhance the sensitivity of the measurement while preserving model independence, an adversarial deep neural network (ADNN) [5] is used to classify signal and background events, while reducing sensitivity to the specific physics models used for signal. The ADNN is designed to be agnostic to the couplings at the HVV vertex of the signal processes.

The ADNN consists of a classifier and an adversary, that are trained in a competitive way. The classifier performs signal vs background classification and is trained on a data set that includes signal events generated under SM and BSM physics hypotheses. The adversary aims to infer the underlying physics model of a signal event from the second-to-last layer of the classifier. If the classifier internal representation contains too much information on the signal hypothesis, the classifier is penalized by the adversary. Technically, this is implemented by adding a penalty term to the classifier loss function, proportional to the adversary loss. More details on the training procedure can be found in Ref. [5].

Two separate ADNNs are employed to discriminate VBF and ggF signals with a model-independent approach. The VBF-ADNN is trained on samples including the SM hypothesis as well as various BSM scenarios with ACs affecting the HVV vertex, in both pure BSM and mixed SM-BSM configurations. Similarly, the GGH-ADNN targets the ggF signal and is trained on SM and BSM hypotheses, including pure and mixed scenarios. The two ADNNs have been implemented using the KERAS and TENSORFLOW libraries, whereas the hyperparameters optimization has been done using the OPTUNA package. The training is performed in the inclusive SR detailed in the previous section, using MC simulations of the signal process generated under both the SM and the AC assumptions, alongside the main background contributions. The input features include observables related to the dilepton and dijet systems, event-level quantities and matrix-element-based discriminants.

Figure 1 shows the output distributions of the VBF-ADNN (\mathcal{D}_{VBF}) and GGH-ADNN (\mathcal{D}_{ggF}) for signal and background events. The signal includes all coupling hypotheses used in the training, with the SM case represented by dots and associated uncertainties reflecting the statistical error due to the limited size of the simulated samples. Similar uncertainties affect the BSM signals, as all hypotheses share the same number of training events. The discriminator shapes are consistent across signal models, suggesting limited sensitivity of the classifier to the specific coupling configuration. Both networks show strong separation between signal and background, with binary classification accuracies of 91% for VBF-ADNN and 79% for GGH-ADNN.

4. – Results

The differential production cross sections of the Higgs boson are inferred from the signal strength modifiers, obtained through a simultaneous fit to all bins and categories of signal candidate events, together with the two control regions.

The signal is extracted with three fit strategies: the first extracts the combined VBF and ggF cross section; the second measures the VBF and ggF contributions simultane-

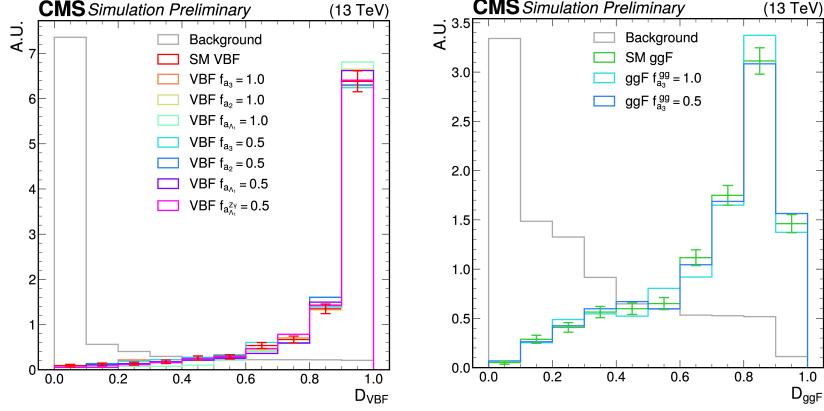


Fig. 1. – Normalized distributions of \mathcal{D}_{VBF} (left) and \mathcal{D}_{ggF} (right), evaluated on signal and background events. The signal is displayed for the different coupling hypotheses, with the SM contribution represented by dots and associated error bars reflecting the statistical uncertainty, and alternative hypotheses in solid lines [3].

ously; and the third determines the VBF-only cross section while fixing the ggF process to the SM prediction. The fit variable chosen for the signal extraction procedure depends on the specific cross section being measured. The VBF differential cross section is measured using the \mathcal{D}_{VBF} distribution. For the other two fit configurations, a two-dimensional variable ($\mathcal{D}_{\text{VBF,ggF}}$) is used, built from the outputs of two adversarial neural networks \mathcal{D}_{VBF} and \mathcal{D}_{ggF} .

The differential cross sections obtained from the all fit configurations are illustrated in Fig. 2. The error bars represent the total uncertainty at the 68% confidence level (CL), with the statistical component being the dominant source.

The model dependence of the measured cross section is quantified as the potential bias arising from variations in the signal model. It reflects the impact of assuming the SM hypothesis when extracting the signal in scenarios that include BSM contributions. The use of the ADNN reduces the associated bias by 30% to 70% across the different $\Delta\Phi_{\text{jj}}$ bins when compared to a non-adversarial method. The remaining model dependence is typically smaller than the total uncertainty and it primarily arises from acceptance effects not addressed by the ADNN.

Following the measurement of the differential cross sections, additional insight into the $\Delta\Phi_{\text{jj}}$ observable is obtained by studying its asymmetry. A nonzero value would indeed indicate CP violation in the Higgs sector. The asymmetry is defined as:

$$(2) \quad A = \frac{N(\Delta\Phi_{\text{jj}} > 0) - N(\Delta\Phi_{\text{jj}} \leq 0)}{N(\Delta\Phi_{\text{jj}} > 0) + N(\Delta\Phi_{\text{jj}} \leq 0)}.$$

and the observed value of the is $A = -0.4^{+0.3}_{-0.3}$, with the total uncertainty corresponding to the 68% CL.

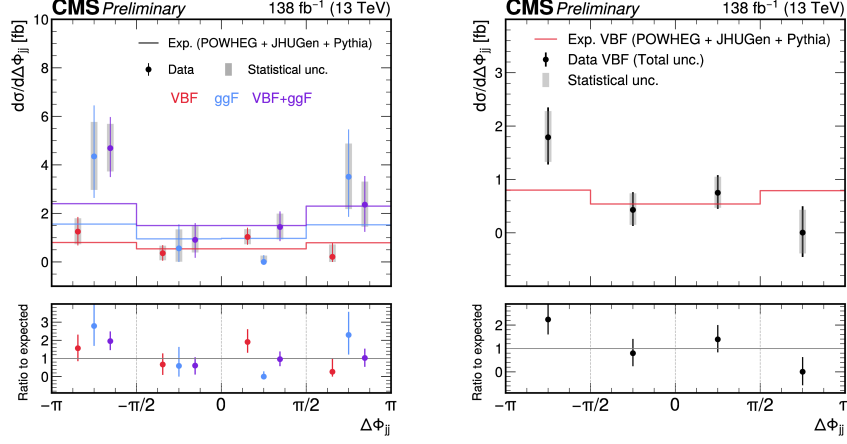


Fig. 2. – Measured differential cross section from fit configuration 1 and 2 (left) and from configuration 3 (right). Colored markers represent the extracted cross section values from data, with error bars showing the combined statistical and systematic uncertainties. The gray bands indicate the statistical uncertainties. The colored histogram corresponds to the expected SM prediction, simulated with POWHEG + JHUGen + Pythia generators. The lower panel displays the ratio of the measured values to the SM expectation [3].

5. – SMEFT interpretation

The differential cross section measurements are interpreted within the SMEFT framework, which extends the SM Lagrangian by including higher-dimensional operators that encode possible effects of the new physics at a high-energy scale $\Lambda \gg 1$ TeV. These operators are built from the SM fields and respect the SM symmetries. They are suppressed by powers of Λ and their strength is determined by dimensionless Wilson coefficients.

In this analysis, only dimension-6 operators are considered, as operators of dimension-5 and -7 typically violate lepton and baryon number conservation. The coefficients c_{HW} , c_{HWB} , c_{HB} , c_{HG} are associated with operators that conserve CP symmetry in the Higgs boson interaction with vector bosons, while $c_{H\tilde{W}}$, $c_{H\tilde{W}B}$, $c_{H\tilde{B}}$, $c_{H\tilde{G}}$ violate it. The coefficients c_{HG} and $c_{H\tilde{G}}$ primarily affect ggF production, whereas the other coefficients influence the VBF process. The coefficients $c_{H\Box}$ and c_{HD} are also considered to probe sensitivity to anomalous corrections in the kinetic terms of the scalar fields.

The strongest constraints were obtained for the CP-even coefficient c_{HW} , which is sensitive to the coupling between the Higgs boson and the W bosons, and for the CP-even coefficient c_{HG} , which probes the interaction between the Higgs boson and gluons. Finally, the differential cross section was re-evaluated using the observed value of the Wilson coefficients and compared with the results from the previous section. The comparison is presented in Fig. 3 for the coefficients c_{HW} , c_{HWB} , c_{HB} and their CP-odd counterparts.

6. – Conclusions

This note presents a model-independent measurement of the Higgs boson differential production cross section in the WW decay channel, with a final state containing two

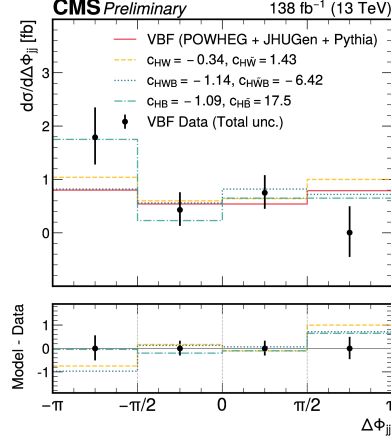


Fig. 3. – Measured differential cross section for VBF production as a function of $\Delta\Phi_{jj}$ (black) compared to various predictions. The predictions include: the SM (red), the ones obtained from the best-fit of Wilson coefficients of c_{HW} , $c_{H\bar{W}}$ (yellow), c_{HWB} , $c_{H\bar{W}B}$ (blue) and c_{HB} , $c_{H\bar{B}}$ (green). The difference between the data and the predictions are displayed in the bottom panel [3].

jets and a different-flavor dilepton pair. The measurement is based on proton-proton collision data recorded by the CMS detector from 2016 to 2018, corresponding to an integrated luminosity of 138 fb^{-1} at a center-of-mass energy of 13 TeV. The production cross section is measured as a function of the azimuthal angle difference between the two jets in the final state. Three different signal extraction strategies are employed to measure the Higgs boson production in association with two jets, targeting both VBF and gluon fusion ggF production mechanisms. No significant deviations from the SM expectations are observed. The measured cross sections are further interpreted in the SMEFT framework to constrain Wilson coefficients. The most stringent constraints are obtained CP-even coefficient c_{HW} , which parametrizes anomalous couplings between the Higgs and W bosons. Similarly, the ggF cross section provides strong sensitivity to the CP-even coefficient c_{HG} , which governs the effective interaction between the Higgs boson and gluons. All results are found to be consistent with SM predictions.

REFERENCES

- [1] CMS Collaboration, The CMS experiment at the CERN LHC, *JINST* **3**, S08004 (2008)
- [2] T. Plehn, D. L. Rainwater and D. Zeppenfeld, Determining the Structure of Higgs Couplings at the LHC, *Phys. Rev. Lett* **88**, 051801 (2002)
- [3] CMS Collaboration, Model-independent measurement of the Higgs boson differential production cross section in association with two jets in the WW decay channel, <https://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/HIG-24-004/index.html>
- [4] B. Grzadkowski, M. Iskrzyński, M. Misiak and J. Rosiek, Dimension-six terms in the Standard Model Lagrangian, *JHEP* **10**, 085 (2010)
- [5] B. Camaiani et al., Model independent measurements of standard model cross sections with domain adaptation. *EPJC* **82**, 921 (2022)