Available on the CERN CDS information server

CMS PAS HIG-17-015

CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch

2017/03/30

Measurement of differential fiducial cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 13\text{TeV}$

The CMS Collaboration

Abstract

A measurement of the integrated and differential fiducial production cross sections for the Higgs boson in the diphoton decay channel at $\sqrt{s} = 13$ TeV is performed using 35.9 fb^{-1} of pp collisions data collected by the CMS experiment at the CERN LHC. Differential cross sections are measured as a function of the diphoton transverse momentum and jet multiplicity. All cross sections are measured within a fiducial phase space defined by the requirements on the photons kinematics, their isolation, and the event topology. The measured cross sections are compared to state of the art theoretical predictions for the Standard Model Higgs bosons. A good agreement between observations and predictions is observed.

1 Introduction

The discovery of the Higgs boson was announced in 2012 by the CMS and ATLAS collaborations [1, 2] based on proton-proton collisions collected at the CERN LHC at the centre of mass energies of 7 and 8 TeV. Since then a vast campaign of measurements has started in order to characterize the newly found particle and compare its properties with those predicted by the Standard Model (SM) of particle physics [3]. The measurement of the $\mathrm{H} \rightarrow \gamma\gamma$ differential production cross sections as a function of kinematic observables allows one to investigate possible deviations from the SM predictions for the dynamics of the production, decay and additional hadronic activity. These investigations are expected to bring important insights on the nature of the Higgs boson and allow to test perturbative QCD predictions in the Higgs sector. Both ATLAS and CMS presented results on the measurement of fiducial and differential cross section for the Higgs boson in the diphoton decay channel at $\sqrt{s} = 8$ TeV [4] [5]. At $\sqrt{s}$ = 13 TeV CMS presented the integrated fiducial cross section with 12.9 fb^{-1} [6] and ATLAS the fiducial and differential cross sections [7].

Despite its small branching ratio of about 0.2% [8] predicted by the SM, the $\mathrm{H} \rightarrow \gamma\gamma$ decay channel provides a clean final-state topology that allows the diphoton invariant mass to be reconstructed with high precision. The dominant sources of background are the irreducible direct diphoton production, and the reducible processes $\mathrm{pp} \rightarrow \gamma + \mathrm{jet}$ and $\mathrm{pp} \rightarrow \mathrm{jet} + \mathrm{jet}$.

With the increasing amount of data collected by the experiments, it becomes increasingly important to have a model-independent assessment of the Higgs boson properties. The standard approach adopted by the CMS collaboration (see [6] for the latest results) in the measurement of $\mathrm{H} \rightarrow \gamma\gamma$ production employs as much information as possible to separate signal from the reducible and irreducible SM backgrounds. This allows to minimize the statistical uncertainties associated with the measurements, but relies on state of the art predictions of the Higgs boson kinematics and on the associated uncertainties. A complementary approach consists in measuring the Higgs boson production as a function of the kinematic properties of the final state. This approach, though not achieving the same statistical power, minimizes the dependence on the assumptions of the SM Higgs boson kinematics. As the statistical uncertainties become smaller and smaller such an approach gains an increasingly larger importance.

In this note we report the measurement of the integrated and differential fiducial cross sections for Higgs boson production at $\sqrt{s}$ =13 TeV in the diphoton decay channel using data corresponding to 35.9 fb^{-1} of proton-proton collisions recorded by the CMS experiment in 2016. The methods used in this note follow closely what was developed for the $\mathrm{H} \rightarrow \gamma\gamma$ differential measurements at 8 TeV [5]. For each bin of the differential variables, the signal is extracted by fitting for a small signal peak on top of the steeply falling spectrum of the diphoton mass distribution. To improve the analysis sensitivity, the selected events are categorized using a diphoton mass resolution estimator, which was decorrelated from the diphoton mass distribution and allows for a smooth background description. No attempt is made to further classify the events according to their production mechanisms, therefore the differential cross sections are mostly sensitive to the gluon fusion. Both integrated and differential cross sections are measured within a fiducial phase space defined by requirements on the photons kinematics, their isolation, and event topology. Differential cross sections are measured as a function of the diphoton system transverse momentum, and as a function of the number of additional hadronic jets in the event.

The note is organized as follows. After a brief description of the CMS detector in Section 2 and of the data and simulated samples in Section 3, the event reconstruction is presented in Section 4, the mass resolution estimator in Section 5 and the event selection in Section 6. The signal extraction, together with the signal and background modeling, the statistical methodology and the unfolding procedure are described in Section 7. Systematic uncertainties are reported in Section 8 and finally the results are presented Section 9.

2 The CMS detector

The central feature of the CMS detector is a superconducting solenoid, 13 m in length and 6 m in diameter, which provides an axial magnetic field of 3.8 T. The core of the solenoid is instrumented with both the tracker and the calorimetry. The steel return yoke outside the solenoid is instrumented with gas detectors used to reconstruct and identify muons. Charged particle trajectories are measured by the silicon pixel and strip tracker, with full azimuthal coverage within $|\eta| < 2.5$, where the pseudorapidity η is defined as $\eta = -\ln[\tan(\theta/2)]$, with θ being the polar angle of the trajectory of the particle with respect to the counterclockwise beam direction. A lead-tungstate crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL) surround the tracking volume and cover the region $|\eta| < 3$. The ECAL barrel extends to $|\eta| < 1.479$ while the ECAL endcaps cover the region $1.479 < |\eta| < 3.0$. A lead/silicon-strip preshower detector is located in front of the ECAL endcap in the region $1.653 < |\eta| < 2.6$. The preshower detector includes two planes of silicon sensors measuring the x and y coordinates of the impinging particles. A steel/quartz-fibre Čerenkov forward calorimeter extends the calorimetric coverage to $|\eta| < 5.0$. In the region $|\eta| < 1.74$, the HCAL cells have widths of 0.087 in both pseudorapidity and azimuth (ϕ). In the (η, ϕ) plane, and for $|\eta| < 1.48$, the HCAL cells map on to 5×5 ECAL crystal arrays to form calorimeter towers projecting radially outwards from points slightly offset from the nominal interaction point. In the endcap, the ECAL arrays matching the HCAL cells contain fewer crystals. A detailed description of the CMS detector can be found elsewhere [9].

3 Data and Monte Carlo samples

The events used in the analysis were selected by a diphoton trigger with asymmetric transverse energy thresholds 30 GeV on the leading p_T photon, 18 GeV on the trailing p_T one, a minimum invariant diphoton mass of 90 GeV, and loose requirements on the calorimetric isolation of the photon candidates and on the shape of its electromagnetic shower.

The Monte Carlo (MC) simulation of the detector response is based on a model implemented in GEANT4 [10]. Simulated events include simulation of the multiple pp interactions taking place in each bunch crossing and are weighted to reproduce the distribution of the number of interactions in data.

The signal samples are generated with MADGRAPH_aMC@NLO [11] and POHWHEG [12] for the four main Higgs boson production modes: gluon-gluon fusion (ggH), vector boson fusion (VBF), associated production with a W/Z boson (VH), and associated production with a $t\bar{t}$ pair (ttH). The parton level samples are interfaced to PYTHIA8 [13] with CUETP8M1 [14] tune, for parton showering and hadronization. The SM Higgs boson cross sections and branching fractions used are taken from the LHC Higgs Cross Section Working Group recommendations [15]. Simulated samples of $Z \to e^+e^-$, $Z \to \mu^+\mu^-$, and $Z \to \mu^+\mu^-\gamma$ events are generated with MADGRAPH_aMC@NLO and used for comparison with data and for the derivation of energy scale and resolution corrections.

Simulated background samples are used for training of multivariate discriminants, and to define selection and classification criteria. The SM production of two prompt photons ($\gamma\gamma$) is simulated using SHERPA1.4.2 [16]. The remaining processes, where one of the photon candidates arises from misidentified jet fragments, are simulated using PYTHIA alone.

4 Event reconstruction

Photon candidates are reconstructed from clusters of the energy deposited in the ECAL and merged into superclusters. The reconstruction algorithm of photon clusters allows almost complete recovery of the energy also of photons that convert to an electron positron pair in the material in front of the ECAL. A detailed description of the algorithm can be found in Ref. [17].

In order to obtain the best energy resolution, the calorimeter signals are calibrated and corrected for several detector effects [18]. Calibration of the ECAL uses photons from $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ and electrons from $W \to e\nu$, and $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ decays. Changes in the transparency of the ECAL crystals due to irradiation during the LHC running periods and their subsequent recovery are monitored continuously and corrected for, using light injected from a laser and LED system [18]. In the barrel section of the ECAL, an energy resolution of about 1% is achieved for unconverted or late-converting photons in the tens of GeV energy range. The remaining barrel photons have a resolution of about 1.3% up to a pseudorapidity of $|\eta| = 1$, rising to about 2.5% at $|\eta| = 1.4$. In the endcaps, the resolution of unconverted or late-converting photons is about 2.5%, while the remaining endcap photons have a resolution between 3 and 4% [19].

The containment of the shower in the clustered crystals, the shower losses for photons that convert in the material upstream of the calorimeter, and the effects of pileup, are corrected using a multivariate regression technique that is trained on simulated events using as input shower shape variables and variables related to the photon position in the detector. The photon energy response distribution is parameterized by a function with a Gaussian core and two power law tails, an extended form of the Crystal Ball function [20]. The regression provides a per-photon estimate of the parameters of the function, and therefore a prediction of the distribution of the ratio of true energy to uncorrected supercluster energy. The most probable value of this distribution is taken as the corrected photon energy. The width of the Gaussian core is further used as a per-photon estimator of the relative energy resolution σ_E/E.

Because no ionization energy is deposited in the tracker by photons, the assignment of the diphoton candidate to a vertex can only be done indirectly exploiting the the properties of each reconstructed primary vertex. Three discriminating variables are calculated for each reconstructed vertex: the sum of the squared transverse momenta of the charged particle tracks associated with the vertex, and two variables that quantify the vector and scalar balance of p_{T} between the diphoton system and the charged particle tracks associated with the vertex. In addition, if either photon is associated with any charged particle track that has been identified as resulting from conversion, also the conversion information is used. The variables are used as the inputs to a multivariate classifier based on a boosted decision tree (BDT) to choose the reconstructed vertex to be associated with the diphoton system. The average vertex finding efficiency of this algorithm is 80%.

A photon identification algorithm is used to separate prompt photons from photon candidates resulting from the misidentification of jet fragments [21]. The algorithm is implemented with a BDT trained on Monte Carlo events. The input variables of the BDT are: the pseudorapidity and energy of the supercluster corresponding to the reconstructed photon, several variables characterizing the shape of the electromagnetic shower and the isolation energies computed with the particle flow (PF) algorithm [22, 23]. The inclusion of the photon pseudorapidity and energy in the list of training inputs, results into a flat photon identification efficiency as a function of η and p_{T}, which reduces the model-dependence of the corrections needed for the final comparisons with theoretical predictions.

Jets are reconstructed within the pseudo-rapidity region $|\eta| < 2.5$, using the anti-kT [24] algorithm with a distance parameter of 0.4 on particles identified by the PF algorithm, excluding charged candidates associated with a vertex other than the selected vertex for the event. Jet energy corrections account in particular for pileup, and are obtained from simulation. In order to include residual data-simulation corrections, they are calibrated with in situ measurements using the energy balance studied in dijet and γ/Z+jet events. The jet momentum scale is found to be within 5-10% of the true jet momentum over the whole spectrum and detector acceptance. The jet energy resolution is typically 15% and 8% at 10 and 100 GeV, respectively. Jets are selected if they fail the pileup identification criteria [25] and have $\mathrm{p_T} > 30\,\mathrm{GeV}$. The minimum distance between photons and jets is required to be $\Delta R(\gamma,\mathrm{jet}) = \sqrt{\Delta\eta^2(\gamma,\mathrm{jet}) + \Delta\phi^2(\gamma,\mathrm{jet})} > 0.4$, where $\Delta\eta(\gamma,\mathrm{jet})$ and $\Delta\phi(\gamma,\mathrm{jet})$ are the pseudorapidity and azimuthal angle differences between photons and jets, to minimize photon energy depositions into jets.

5 Mass resolution estimator

The selected photon pairs are categorized according to their estimated relative mass resolution $\sigma_{\mathrm{m}}/\mathrm{m}$. For the typical energy range of the photons used in this analysis, the energy resolution estimator depends on the energy itself because of the stochastic and noise terms in the energy resolution of the electromagnetic calorimeter. The nature of these two terms is such that the energy resolution improves at higher energy. This dependence is propagated to the mass resolution estimator, which is thus dependent on the mass of the diphoton pair. An event categorization simply based on such a variable would distort the mass distribution of the events falling in the different categories.

To avoid such an effect, a new mass resolution estimator, $\sigma_{\mathrm{m}}/\mathrm{m}|_{\mathrm{decorr}}$, is built by decorrelating the diphoton mass from the mass resolution estimator $\sigma_{\mathrm{m}}/\mathrm{m}$. The decorrelation employs the cumulative distribution functions of the $\sigma_{\mathrm{m}}/\mathrm{m}$ estimator, conditional on the diphoton mass m. The transformation $\mathrm{y(x)} = \int_{\mathrm{x_{min}}}^{\mathrm{x}} \mathrm{f(x')dx'}$ maps any continuous random variable x into a uniformly distributed one. The decorrelation procedure therefore constructs a mass-dependent transformation

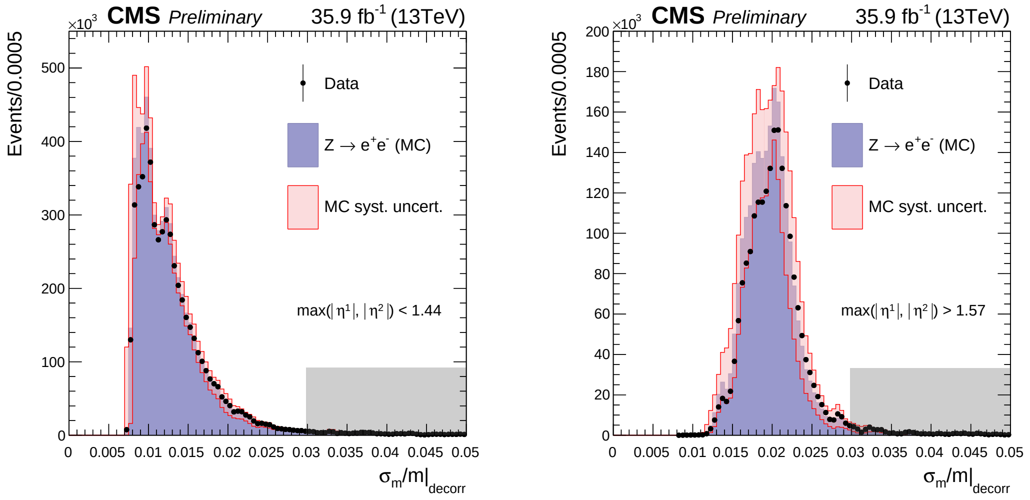
$$\mathrm{y}(\sigma_{\mathrm{m}}/\mathrm{m}|\mathrm{m}) = \int_0^{\sigma_{\mathrm{m}}/\mathrm{m}} f(\sigma_{\mathrm{m}}/\mathrm{m}'|\mathrm{m})\,d\sigma_{\mathrm{m}}/\mathrm{m}'$$

where the conditional distribution of $\sigma_{\mathrm{m}}/\mathrm{m}$, $f(\sigma_{\mathrm{m}}/\mathrm{m}|\mathrm{m})$, is constructed by sorting the values of $\sigma_{\mathrm{m}}/\mathrm{m}$ in bins of m.

The construction of such a transformation in bins of the diphoton mass ensures a flat mass resolution estimator in each bin, uniformly distributed and thus no longer correlated with the mass. The inverse cumulative distribution at a given mass can be used to further transform the mass resolution estimator into a quantity interpretable as a physical mass resolution.

The decorrelation transformation is applied to the simulated $\gamma\gamma$ sample, using $m_{\gamma\gamma} = 125$ GeV as a reference point for the inverse transformation. The transformation is derived separately for events where both photons are reconstructed in the barrel region and for events where at least one photon is reconstructed in the endcap region.

The modelling of the decorrelated mass resolution estimator is studied with $\mathrm{Z} \to \mathrm{e^+e^-}$ events where electrons are reconstructed as photons. The per-photon resolution estimate is affected by the imperfect modeling of the electromagnetic shower shape in the simulation. To reduce the disagreement [26], the per-photon resolution estimate is recomputed using as input simulated



(a) (b)

Figure 1: Data-simulation comparison for the decorrelated mass resolution estimator $\sigma_m/m|_{decorr}$, along with the impact of the systematic uncertainty on the σ_E/E (red band), for events with both photons in the ECAL barrel (left) and the remainder of the events. Events in region depicted in grey are discarded and not used in the final analysis.

shower shapes corrected to match the ones observed in data. A systematic uncertainty is assigned multiplying and dividing the value of σ_E/E by 1.05 for each photon candidate, to cover the residual discrepancy. Figure 1 shows the data-simulation comparison for the decorrelated mass resolution estimator $\sigma_m/m|_{decorr}$, along with the impact of the systematic uncertainty on the σ_E/E (red band).

6 Event selection and categorization

Each photon of the candidate pair entering the analysis is required to be within $|\eta| < 2.5$, excluding the region $1.4442 < |\eta| < 1.566$, and to satisfy selection criteria similar to, but slightly more stringent than, the trigger requirements, based on transverse momentum, isolation and shower shape variables. Furthermore each photon needs to fulfill a requirement based on the output of the photon identification classifier.

The analysis aims at providing the measurement of the production cross section of the Higgs boson in a fiducial phase space, to be compared with theoretical predictions. The selection criteria in data are necessarily defined at reconstruction level, while the fiducial phase space, for which theory predictions are computed, is defined without considering the detector response. Because of the finite detector resolution, the two definitions do not exactly coincide, and for this reason events fulfilling the event selection criteria at reconstruction level can originate from either inside or outside the corresponding fiducial region. To minimize the number of events entering the latter category, the selection criteria imposed at reconstruction level were defined to mimic the definition of the fiducial phase space as closely as possible.

The analysis fiducial phase space is defined requiring the leading (subleading) photon transverse momentum scaled by the invariant mass of the diphoton candidate ($p_T/m_{\gamma\gamma}$) to be greater than 1/3 (1/4), and requiring the absolute pseudorapidity of both photons to be within 2.5. In addition, a requirement on the maximum generator level total hadronic energy contained in a cone of radius $\Delta R = 0.3$ around each photon candidate is required to be less than 10 GeV.

Events are split into categories of $\sigma_\mathrm{m}/\mathrm{m}|_\mathrm{decorr}$ to maximize the analysis sensitivity to the SM Higgs boson. The positions of the $\sigma_\mathrm{m}/\mathrm{m}|_\mathrm{decorr}$ boundaries are optimized simultaneously with the lower bound on the output of the photon identification classifier. Three categories are adequate to saturate the maximum sensitivity achievable with this method on the present dataset. The efficiency associated to the photon identification selection is roughly 80% for signal events in the fiducial phase space. The categories obtained from the optimization process correspond approximately to the configurations where both photons are reconstructed in the central barrel ($|\eta| < 1$) for the first category, both photons are reconstructed in the barrel with at least one falling outside the central barrel for the second category and at least one photon reconstructed in the endcap regions for the last category.

The identification and trigger efficiencies are measured using data events containing a Z boson decaying to a pair of electrons, or to a pair of electrons or muons in association with a photon [19]. After applying data driven corrections to the input of the photon identification classifier, the efficiencies measured in data are found to be 3%(5%) lower than predicted for photons in the barrel (endcap) regions with $R_9 < 0.85$ (0.9). A correction factor is applied to simulated events to take this into account. For the remaining photons the predicted efficiencies are compatible with the ones measured in data.

7 Statistical analysis

In order to measure the differential production cross section, the events fulfilling the selection criteria are further split into two set of bins, based on the diphoton transverse momentum, and the number of additional hadronic jets, respectively. In the case of the diphoton transverse momentum, 8 bins are used $\{[0\,,15), [15\,,30), [30\,,45), [45\,,85), [85\,,125), [125, 200), [200, 350), [350, \infty)\}$ GeV, while 5 are used in the case of the number of jets {n=0, n=1, n=2, n=3, n≥4}. For each of the kinematic bins, the signal is extracted splitting the events in three categories according to the $\sigma_\mathrm{m}/\mathrm{m}|_\mathrm{decorr}$ variable as described in the previous section.

The signal production cross section is extracted through a simultaneous extended maximum likelihood fit to the diphoton invariant mass spectrum in all the analysis categories. The complete likelihood reads as:

$$\begin{aligned}
&\mathcal{L}(\mathrm{data}|\vec{\Delta\sigma}, \vec{n_{bkg}}, \vec{\theta_S}, \vec{\theta_B}) = \\
&\prod_{i=1}^{n_{cat}}\prod_{j=1}^{n_b}\prod_{l=1}^{n_{m_{\gamma\gamma}}}\left(\sum_{k=1}^{n_b} L\Delta\sigma_k K_k^{ij}(\vec{\theta_S}) S_k^{ij}(m^l_{\gamma\gamma}|\vec{\theta_S}) + n^{ij}_{OOA} S^{ij}_{OOA}(m^l_{\gamma\gamma}|\vec{\theta_S}) + n^{ij}_{bkg} B^{ij}(m^l_{\gamma\gamma}|\vec{\theta_B})\right)^{n^{lij}_{ev}} \cdot \qquad (1)\\
&\mathrm{Pois}(n^{ij}_{ev}|n^{ij}_{sig} + n^{ij}_{bkg}) \cdot Pdf(\vec{\theta_S}) Pdf(\vec{\theta_B})
\end{aligned}$$

where:

- L is the total integrated luminosity analyzed;
- $\vec{\Delta\sigma} = (\Delta\sigma_1 ... \Delta\sigma_{n_b})$ is the vector of fiducial cross sections being measured;
- K_k^{ij} are the response matrices, which represent the probability that an event in the k-th kinematic bin at generator level is reconstructed in the ij-th reconstruction-level category (with the index i running on the $\sigma_\mathrm{m}/\mathrm{m}|_\mathrm{decorr}$ categories and the index j running on the kinematic bin).

- the functions S_k^{ij} and B^{ij} are the signal and background distributions in $m_{\gamma\gamma}$, which will be described in the next section;
- n_{ev}^{ij}, n_{sig}^{ij}, n_{bkg}^{ij} are the number of observed, signal and background events in the ij-th reconstruction-level category respectively;
- the terms $n_{OOA}^{ij} S_{OOA}^{ij}$ represent the contributions of the Higgs boson signal originating outside of the fiducial region. The contribution of the out of acceptance Higgs boson signal is estimated from simulations to be roughly 1% of the total expected SM signal;
- the parameters $\vec{\theta}_S$ and $\vec{\theta}_B$ are the nuisance parameters associated with the signal and background model respectively.

The unfolding to particle-level cross sections is achieved with these definitions by extracting the vector $\vec{\Delta\sigma}$ directly from the likelihood fit. The uncertainties and the correlation matrices are obtained from the test statistics $q(\Delta\vec{\sigma})$, which is defined as reported below and it is asymptotically distributed as a χ^2 with n_b degrees of freedom[27]:

$$q(\Delta\vec{\sigma}) = -2 \cdot \log \left(\frac{\mathcal{L}(\Delta\vec{\sigma}|\hat{\vec{\theta}}_{\Delta\vec{\sigma}})}{\mathcal{L}(\Delta\hat{\vec{\sigma}}|\hat{\vec{\theta}})} \right) \tag{2}$$

where $\vec{\theta} = (n_{bkg}, \vec{\theta}_S, \vec{\theta}_B)$. The notation $\hat{y}$ represents the best fit estimate of y, while $\hat{y}_x$ stands for the best fit estimate of y, conditionally on the value of x. The nuisance parameters, including the Higgs boson mass, are profiled in the fit across all the bins.

7.1 Signal model

For a given observable, a parametric signal model is constructed separately for each generator-level bin (adding an extra bin collecting the out-of-acceptance events), for each reconstruction-level bin and for each of the three bins in $\sigma_{\mathrm{m}}/\mathrm{m}|_{\mathrm{decorr}}$. The model is built as a fit to a sum of Gaussian distributions of the simulated invariant mass shape, after applying the already mentioned data-simulation corrections, for three values of $m_{\mathrm{H}} \in \{120, 125, 130\}$ GeV. Signal models for intermediate values of m_{H} are obtained by linear interpolation of the fitted parameters.

7.2 Background model

A background model is built for each of the bins of the differential observable and for each of the three bins in $\sigma_{\mathrm{m}}/\mathrm{m}|_{\mathrm{decorr}}$. A discrete profiling method has been used as initially developed for the $\mathrm{H} \rightarrow \gamma\gamma$ discovery [1]. The background is evaluated by fitting the $m_{\gamma\gamma}$ distribution in data in the range $100 < m_{\gamma\gamma} < 180\,\mathrm{GeV}$, without reference to the MC simulation.

The choice of function used to fit the background in any particular event class is included as a discrete nuisance parameter in the formulation used to extract the result. Exponentials, power-law functions, polynomials in the Bernstein basis, and Laurent polynomials are used to represent $B(m_{\gamma\gamma}|\vec{\theta}_B)$ in Eq.1.

When fitting a signal+background hypothesis to the data, by minimizing the value of twice the negative logarithm of the likelihood, all functions in these families are tried, with a "penalty" term added to account for the number of free parameters in the fit. The penalized likelihood function $\widetilde{\mathcal{L}}_B$ for a single fixed background fitting function B is defined through:

$$-2\ln\widetilde{\mathcal{L}}_B = -2\ln\mathcal{L}_B + N_B, \tag{3}$$

where $\mathcal{L}_B$ is the "unpenalized" likelihood function, and N_B is the number of free parameters in B. When fitting the complete likelihood, the number of degrees of freedom (number of exponentials, number of terms in the series, degree of the polynomial, etc.) is increased until no significant improvement (p value < 0.05 obtained from the F-distribution) occurs in the likelihood between N+1 and N degrees of freedom for the fit to data.

8 Systematic uncertainties

Systematic uncertainties listed in this section are included in the likelihood as nuisance parameters and are profiled during the minimization. Unless specified to the contrary, the sources of uncertainty refer to the individual quantity studied, and not to the final yield. The uncertainty of the present measurement is dominated by statistical uncertainties. The systematic uncertainties considered in this analysis are the following:

- *Integrated luminosity*: the systematics uncertainty is estimated from data to be 2.6%.
- *Trigger efficiency*: the trigger efficiency is measured from $Z \to e^+e^-$ events using the tag-and-probe technique; the size of the uncertainty is about 1%.
- *Photon selection*: the systematic uncertainty is taken as the uncertainty on the ratio between the efficiency measured in data and in simulation; it ranges from 0.3% to 3.2% according to the photon category and results in an event yield variation from 0.7% to 4.0% depending on the category.
- *Vertex finding efficiency*: the largest contribution to the uncertainty comes from the modeling of the underlying event, plus the uncertainty on the measurement of the corresponding to the ratio of data and simulation obtained using $Z\to \mu^+\mu^-$ events. It is handled as an additional nuisance parameter built into the signal model which allows the fraction of events in the right vertex/wrong vertex scenario to change. The size of the uncertainty of the vertex selection efficiency is 1.5%.
- *Energy scale and resolution*: scale and resolutions are studied with electrons from $Z \to e^+e^-$ and then applied to photons. The main source of systematic uncertainty is the different interactions of electrons and photons with material upstream the ECAL. Uncertainties are assessed by changing the R_9 distribution, the regression training (using electrons instead of photons) and the electron selection used to derive the corrections. The uncertainty on the additional energy smearing is assigned propagating the uncertainties on the various $|\eta|$ and R_9 bins to the Higgs boson signal phase space. In both cases dedicated nuisance parameters are included as additional systematic terms in the signal model and amount to less than about 0.5% depending on the photon category.
- *Photon identification BDT score*: in order to cover the observed discrepancies between data and simulation, the uncertainty on the signal yields in the different categories of the analysis is estimated conservatively by propagating the uncertainty on the inputs to the final photon ID shape.
- *Per photon energy resolution estimate*: it is parametrized as a rescaling of the resolution estimate by ±5% about its nominal value.
- *Jet energy scale and smearing corrections*: this uncertainty is implemented as migration between jet bins. The size of such migrations is in the 10-20% range, depending on the jet bin.

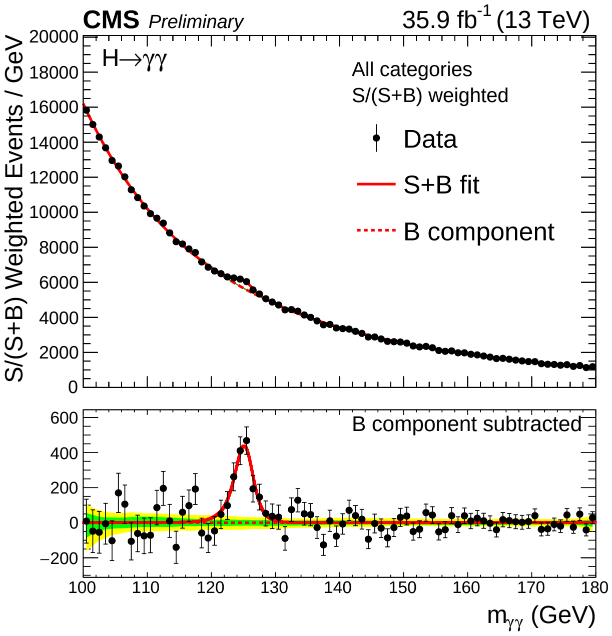


Figure 2: Diphoton mass spectrum weighted by the ratio $S/(S+B)$ in each event class, together with the background subtracted weighted mass spectrum.

- *Background modeling*: the choice of background parametrization is handled using the discrete profiling method. This automatically leads to an uncertainty on the choice of background function and no additional systematic uncertainty needs to be added.
- *Parton density functions (PDF) uncertainties*: the effect of the uncertainty from the choice of PDF is assessed by estimating the relative yield variation in each bin of the observable variables and category, after re-weighting the events of the simulated signal sample. The re-weighting is done according to PDF4LHC15 combined PDF set and NNPDF30 [28] using the MC2hessian procedure [29]. The category migrations are found to be less than 0.3%.
- *QCD scale uncertainty*: related to varying the renormalization and factorization scales. The effect is found to be negligible

9 Results

The data and the best-fit signal-plus-background fit, for the measurement of the inclusive fiducial cross section, are shown in Figure 2 .

The fiducial cross section is measured to be:

$$\hat{\sigma}_{\text{fiducial}} = 84 \pm 11 \text{ (stat) } \pm 7 \text{ (syst) fb} = 84^{+13}_{-12} \text{ (stat+syst) fb}$$

The corresponding likelihood scan is shown in Fig. 3, together with the predicted cross section from theory. There, and in differential cross sections measurements, the value of the Higgs boson mass is treated as a nuisance parameter and profiled in the likelihood maximizations. The value of the profiled mass is compatible with the world average [30].

The theoretical prediction was computed using Monte Carlo events generated with MADGRAPH_aMC@NLO, where each of the Higgs production mechanisms was normalized to the

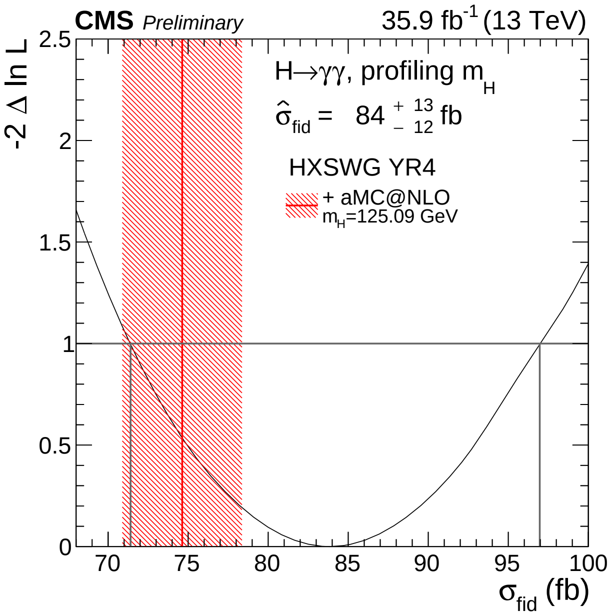


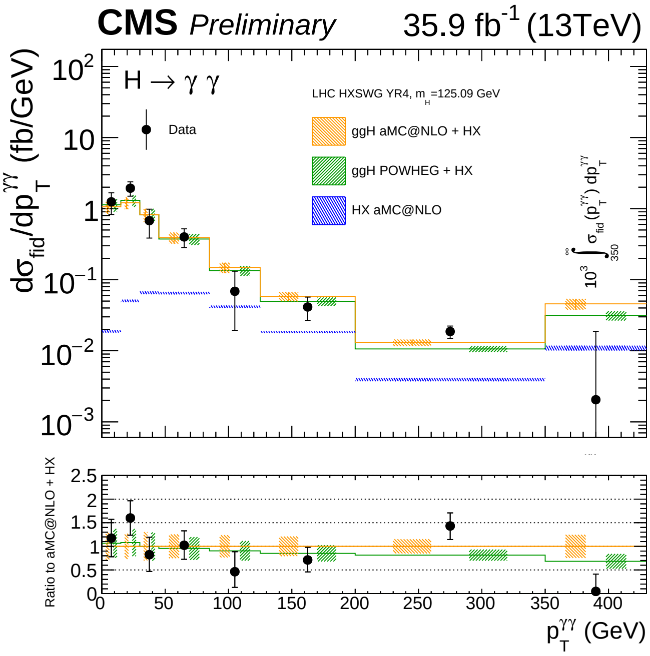
Figure 3: Likelihood scan (black curve) for the fiducial cross section where the value of the SM Higgs boson mass is profiled in the fit. The measurement is compared to the theoretical prediction (in red) and shows good agreement within uncertainties.

predictions from Ref. [15]. In practice, the Monte Carlo events were used to compute the fiducial region acceptance predicted for a SM Higgs boson with a mass of 125.09 GeV and such number was then multiplied by the corresponding total cross section quoted in Ref. [15]. The fiducial region acceptance was estimated to be 0.60 for the SM Higgs boson. Such quantity has a small dependence on the assumed Higgs boson production mechanism, and it amounts to 0.60, 0.60, 0.52, and 0.52 for ggH, VBF, VH, and ttH production, respectively. The associated scale uncertainty was estimated by varying independently the renormalization and factorization scales used in the calculation by a factor of 2 upwards and downwards, excluding the combinations (1/2,2) and (2,1/2), and it amounts to roughly 1% of the acceptance value. If POWHEG is used to generate ggH events instead of MADGRAPH_aMC@NLO, the value of the estimated acceptance changes by roughly 1%.

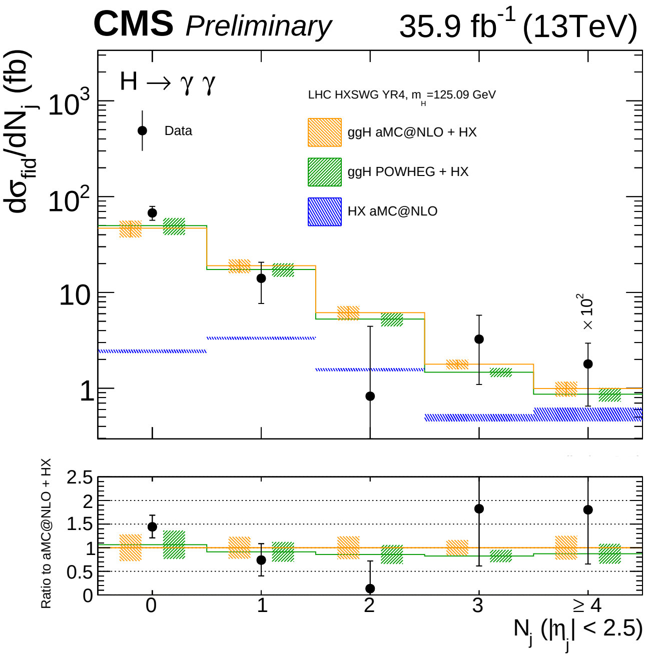
The resulting theoretical prediction was found to be $\sigma^{theory}_{\mathrm{fiducial}} = 75^{+4}_{-4}$fb agreeing with the measured value within uncertainties.

The differential cross sections as a function of the diphoton transverse momentum and the jet multiplicity are reported in Figure 4, together with the corresponding theoretical predictions. Two sets of predictions are shown in each plot. For the first, shown in orange, MADGRAPH_aMC@NLO was used to simulate all the Higgs production processes. The second, shown in green, and was obtained using POWHEG-generated ggH events, while taking other production mechanisms from MADGRAPH_aMC@NLO. The plots show in blue the sum of the contributions from VBF, VH and ttH (labeled HX). The uncertainties on the theoretical predictions were estimated by summing in quadrature the uncertainties from Ref. [15] and the scale uncertainties associated with the MC predictions, computed from the envelope of the predictions obtained varying the renormalization and factorization scales.

The measurements of the fiducial cross sections show a general good agreement with the SM predictions, both inclusively and as a function of the diphoton transverse momentum and the



(a) $p_{T,\gamma\gamma}$ differential cross-section



(b) N_{jets} differential cross-section

Figure 4: Measured H $\rightarrow\gamma\gamma$ differential cross-section (black points) for (a) $p_{T,\gamma\gamma}$, (b) N_{jets}. The measurements are compared to the theoretical predictions, combining the Higgs boson cross sections and branching fraction as in the LHC Higgs Cross Section Working Group [15] with two different generators for the gluon-gluon fusion process: MADGRAPH_aMC@NLO (in orange) and POWHEG (in green). The sum of the contributions from VBF, VH and ttH processes, labeled as HX, is generated using MADGRAPH_aMC@NLO and is shown in blue in the plot.

jet multiplicity.

10 Summary

The measurement of the integrated and differential fiducial production cross sections for the Higgs boson in the diphoton decay channel has been performed using 35.9 fb^{-1} of proton-proton collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV. The measurement of the differential cross section are reported as a function of the Higgs boson transverse momentum and the jet multiplicity. The measurements are performed for isolated photons in the fiducial region defined requiring both photons within $|\eta^{\gamma}_{1,2}| < 2.5$ and $p_{\mathrm{T}}/m_{\gamma\gamma} > 1/3(1/4)$ for the leading (subleading) photon. In this fiducial region the cross section is measured to be $\hat{\sigma}_{\text{fiducial}} = 84 \pm 11$ (stat) ± 7 (syst) fb $= 84^{+13}_{-12}$ (stat+syst) fb.
The presented measurements are the most precise to date and show good agreement with the Standard Model predictions.

References

[1] 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, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[2] 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, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[3] "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *Journal of High Energy Physics* **2016** (2016), no. 8, 45, `doi:10.1007/JHEP08(2016)045`.

[4] "Measurements of fiducial and differential cross sections for Higgs boson production in the diphoton decay channel at $\sqrt{s} = 8$ TeV with ATLAS", *Journal of High Energy Physics* **2014** (2014), no. 9, 112, `doi:10.1007/JHEP09(2014)112`.

[5] "Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV", *The European Physical Journal C* **76** (2016), no. 1, 13, `doi:10.1140/epjc/s10052-015-3853-3`.

[6] "Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s} = 13$ TeV in pp collisions at CMS", technical report.

[7] ATLAS Collaboration, "Measurement of fiducial, differential and production cross sections in the $\mathrm{H} \to \gamma\gamma$ decay channel with 13.3fb^1 of 13 TeV proton-proton collision data with the ATLAS detector", Technical Report ATLAS-CONF-2016-067, CERN, Geneva, Aug, 2016.

[8] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs Properties", CERN Report CERN-2013-004, 2013. `doi:10.5170/CERN-2013-004`, `arXiv:1307.1347`.

[9] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[10] GEANT4 Collaboration, "GEANT4 - a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[11] J. Alwall et al, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", (2014). `arXiv:1405.0301`.

[12] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, `doi:10.1088/1126-6708/2007/11/070`, `arXiv:0709.2092`.

[13] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput.Phys.Commun.* **178** (2008) 852–867, `doi:10.1016/j.cpc.2008.01.036`, `arXiv:0710.3820`.

[14] P. Skands et al., "Tuning PYTHIA 8.1: the Monash 2013 Tune", (2014). `arXiv:1404.5630`.

[15] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, (2016). `arXiv:1610.07922`.

[16] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* **02** (2009) 007, `doi:10.1088/1126-6708/2009/02/007`, `arXiv:0811.4622`.

[17] “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at s=8TeV”, *Journal of Instrumentation* **10** (2015), no. 06, P06005.

[18] CMS Collaboration, “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$ = 7 TeV”, *JINST* **8** (2013) P09009, `doi:10.1088/1748-0221/8/09/P09009`, `arXiv:1306.2016`.

[19] CMS Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P08010, `doi:10.1088/1748-0221/10/08/P08010`, `arXiv:1502.02702`.

[20] M. Oreglia, “A study of the reactions $\psi' \to \gamma\gamma\psi$”. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236.

[21] CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, `arXiv:1407.0558`.

[22] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[23] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[24] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[25] CMS Collaboration, “Pileup Jet Identification”, CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013.

[26] CMS Collaboration Collaboration, “Electron and photon performance in CMS with the full 2016 data sample.”,.

[27] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1, `doi:10.1140/epjc/s10052-011-1554-0`, `arXiv:1007.1727`.

[28] F. Demartin et al., “The impact of PDF and alphas uncertainties on Higgs Production in gluon fusion at hadron colliders”, *Phys. Rev.* **D82** (2010) 014002, `doi:10.1103/PhysRevD.82.014002`, `arXiv:1004.0962`.

[29] S. Carrazza et al., “An Unbiased Hessian Representation for Monte Carlo PDFs”, *Eur. Phys. J.* **C75** (2015), no. 8, 369, `doi:10.1140/epjc/s10052-015-3590-7`, `arXiv:1505.06736`.

[30] ATLAS Collaboration and CMS Collaboration Collaboration, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (May, 2015) 191803, `doi:10.1103/PhysRevLett.114.191803`.