Chapter 4

Event categorisation

4.1 Introduction

The basic experimental method for the observation of the $H \to \gamma\gamma$ decay is to search for a resonance in the continuous SM diphoton invariant mass spectrum. The sensitivity of this method is enhanced by the categorisation of events by their expected mass resolution. Furthermore, additional particles in the final state can be explointed to gain information about the process by which the Higgs boson was produced in each event. This allows the measurement of the cross-section of each production mode individually. This information can be used to probe the strength of the Higgs boson's interaction with the individual particles involved in each mode.

To first order, the most common production mode, ggH, produces a Higgs boson in isolation. This leaves only the photons resulting from $H \to \gamma\gamma$ decay in the final state. The other production modes (VBF, VH and ttH) produce the Higgs boson with additional particles. The VBF mode has two quarks in the final state, which subsequently hadronize to form jets. The VH mode produces a Higgs boson in association with a W or Z boson, which subsequently decays to charged leptons, neutrinos or quarks, leading to reconstructed leptons, $\not{E}_{\mathrm{T}}$ and jets. Finally, the ttH mode produces the Higgs boson in association with two top quarks, which decay to bottom quarks and either hadrons or leptons. These distinctive topologies can be used to categorise events. The categorisation benefits from information about the expected mass resolution of the diphoton, which is obtained with the BDT described in Section 4.2. The categories targeting VBF, VH and ttH events are then discussed in Sections 4.3, 4.4 and 4.5. Finally, the inclusive Untagged categories and the categorisation hierarchy are discussed in Sections 4.6 and 4.7.

4.2 Diphoton BDT

The ranking of events by their expect mass resolution or signal-to-background ratio is obtained using a per-diphoton BDT, which is hereafter referred to as $BDT_{\gamma\gamma}$. The objectives of the $BDT_{\gamma\gamma}$ are to separate signal-like diphotons (produced by the decay of a Higgs boson) from the SM background diphotons, and to assign a high output score to diphotons with a good mass resolution and where both individual photons have high $BDT_{\gamma\text{ID}}$ scores.

Diphotons are composed of pairs of photons and a primary vertex (selected as described in Chapter 3). Before the $BDT_{\gamma\gamma}$ is applied, the diphotons are required to satisfy the following requirements: both photons must pass a loose requirement on their $BDT_{\gamma\text{ID}}$ score, and the p_{T} of the leading (subleading) photon must be above $m_{\gamma\gamma}/3$ ($m_{\gamma\gamma}/4$).

The $BDT_{\gamma\gamma}$ is required to assess diphotons independently of their invariant mass, otherwise the m_{H} value of the training sample would introduce a bias in the output score. The input variables for the $BDT_{\gamma\gamma}$ are therefore chosen to be uncorrelated with the invariant mass of the diphoton system:

- the transverse momentum of each photon, divided by $m_{\gamma\gamma}$ to remove the correlation with m_{H};
- the η-position of each photon;
- the $BDT_{\gamma\text{ID}}$ output score of each photon;
- $\cos(\Delta\phi)$, the cosine of the angle between the photons in the ϕ-direction;
- $\sigma_{\gamma\gamma}^{\text{RV}}/\, m_{\gamma\gamma}$, the per-event estimated mass resolution of the diphoton, assuming that the selected vertex is within 1 cm of the true vertex in the z-direction ($\Delta z \leq 1\,\text{cm}$);
- $\sigma_{\gamma\gamma}^{\text{WV}}/\, m_{\gamma\gamma}$, the per-event estimated mass resolution of the diphoton, assuming that the selected vertex is more than 1 cm away from the true vertex in the z-direction ($\Delta z > 1\,\text{cm}$);
- p_{rv}, the per-event estimate of the probability that the correct vertex was chosen, obtained using the $BDT_{\text{VTX PROB}}$ described in Section 3.4.2.

The per-event estimated mass resolutions are calculated from the individual photon energy resolution estimates, labelled $\sigma_{\gamma_1}^{E}/E_{\gamma_1}$ and $\sigma_{\gamma_2}^{E}/E_{\gamma_2}$, obtained from the semiparametric regression $BDT_{\gamma\text{E}}$ described in Section 3.3.5. If $\Delta z \leq 1\,\text{cm}$, the dominant contributions to the uncertainty on the mass resolution are the energy resolutions of each photon.

Assuming Gaussian resolution functions, $\sigma_{\gamma\gamma}^{\mathrm{RV}}/\,m_{\gamma\gamma}$ can be obtained by simply adding the individual relative photon energy resolutions in quadrature:

$$\sigma_{\gamma\gamma}^{\mathrm{RV}}/\,m_{\gamma\gamma} = \frac{1}{2}\sqrt{(\sigma_{\gamma_1}^{E}/E_{\gamma_1})^2 + (\sigma_{\gamma_2}^{E}/E_{\gamma_2})^2}. \tag{4.1}$$

However, if $\Delta z > 1\,\mathrm{cm}$, the uncertainty on the opening angle contributes significantly to the mass resolution. The effect is modelled by including an additional term which represents the uncertainty on the mass due to the uncertainty on the vertex position, labelled $\sigma_{\gamma\gamma}^{V}$. The distance between the true vertex and the selected vertex follows a Gaussian distribution which has a width equal to the width of the beamspot multiplied by $\sqrt{2}$. Given the spatial positions of the photons, $\sigma_{\gamma\gamma}^{V}$ can therefore be calculated explicitly, and included in the sum in quadrature:

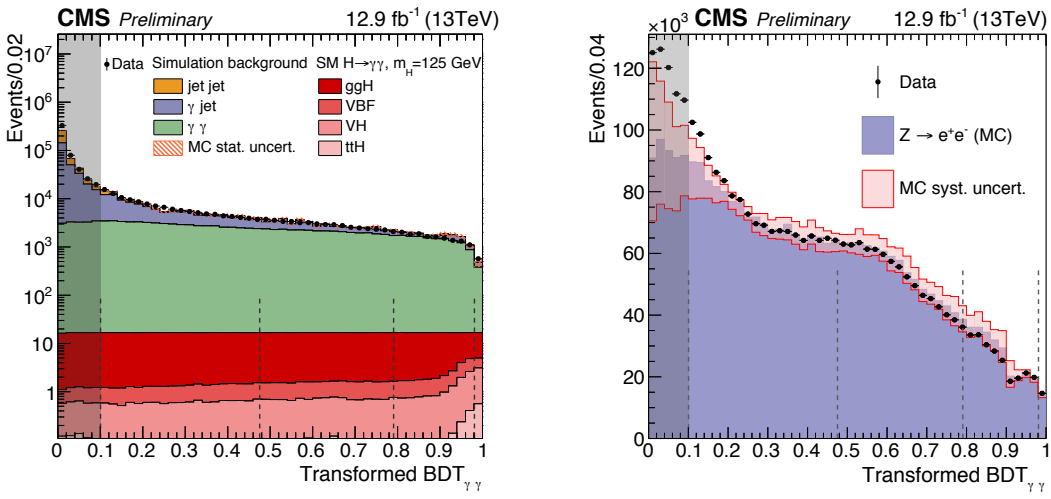
$$\sigma_{\gamma\gamma}^{\mathrm{WV}}/\,m_{\gamma\gamma} = \sqrt{(\,\sigma_{\gamma\gamma}^{\mathrm{RV}}/\,m_{\gamma\gamma})^2 + (\sigma_{\gamma\gamma}^{V}/\,m_{\gamma\gamma})^2}. \tag{4.2}$$

The $BDT_{\gamma\gamma}$ is trained on simulated sampled of signal and background processes. For the signal samples, events from different productions modes (all $m_{\mathrm{H}} = 125\,\mathrm{GeV}$) are mixed according to their cross-section. The signal events used for training are also re-weighted by a factor w given by:

$$w = \frac{p_{rv}}{\sigma_{\gamma\gamma}^{\mathrm{RV}}/\,m_{\gamma\gamma}} + \frac{1-p_{rv}}{\sigma_{\gamma\gamma}^{\mathrm{WV}}/\,m_{\gamma\gamma}}, \tag{4.3}$$

which codifies the fact that the signal-to-background ratio is inversely proportional to the mass resolution. This step ensures that the $BDT_{\gamma\gamma}$ gives a high score to events with good mass resolution. The background for the training is composed of simulated diphotons originating from the irreducible and reducible SM background processes.

Figure 4.1a shows the $BDT_{\gamma\gamma}$ output score for signal and background events in the range $100\,\mathrm{GeV} < m_{\gamma\gamma} < 180\,\mathrm{GeV}$, where a transformation is applied to flatten the signal distribution, so that the efficiency of a particular selection can be easily read. The $BDT_{\gamma\gamma}$ output score is validated using $\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$ events in data and simulation, as can be seen on Figure 4.1b.



(a) **(b)**

Figure 4.1: (a) The transformed $BDT_{\gamma\gamma}$ score for simulated signal and background events in the range $100\,\text{GeV} < m_{\gamma\gamma} < 180\,\text{GeV}$. The transformation flattens the signal distribution. (b) The transformed $BDT_{\gamma\gamma}$ score for $\text{Z} \to \text{e}^{+}\text{e}^{-}$ events in data and simulation, where the electrons are reconstructed as photons. The pink shading represents the systematic uncertainty associated with the $BDT_{\gamma\text{ID}}$ and the $BDT_{\gamma\text{E}}$. For both (a) and (b), the vertical dashed lines represent the boundaries of the Untagged categories described in Section 4.6, while the grey shading represents the area for which diphotons are rejected.

4.3 VBF-tagged categories

The VBF production mode has a cross-section approximately ten times smaller than that of the ggH mode. However, it produces the Higgs boson in association with two quarks, which result in two high-p_{T} jets when reconstructed. This distinctive topology allows the identification of VBF events. For this reason, although VBF events occur much less frequently than ggH events, the most sensitive VBF-tagged category has a signal-to-background ratio comparable to the most sensitive category in the analysis (the Untagged 0 category). Defining VBF-tagged categories therefore significantly improves the overall sensitivity of the analysis.

The VBF-tagged categories are obtained by selecting events which contain two high-p_{T} jets. For each diphoton candidate, the individual jets are obtained as described in Section 3.5.3 after PFCHS with respect to the selected vertex. Candidate VBF-tagged events are required to have at least two jets which pass the pileup removal requirements

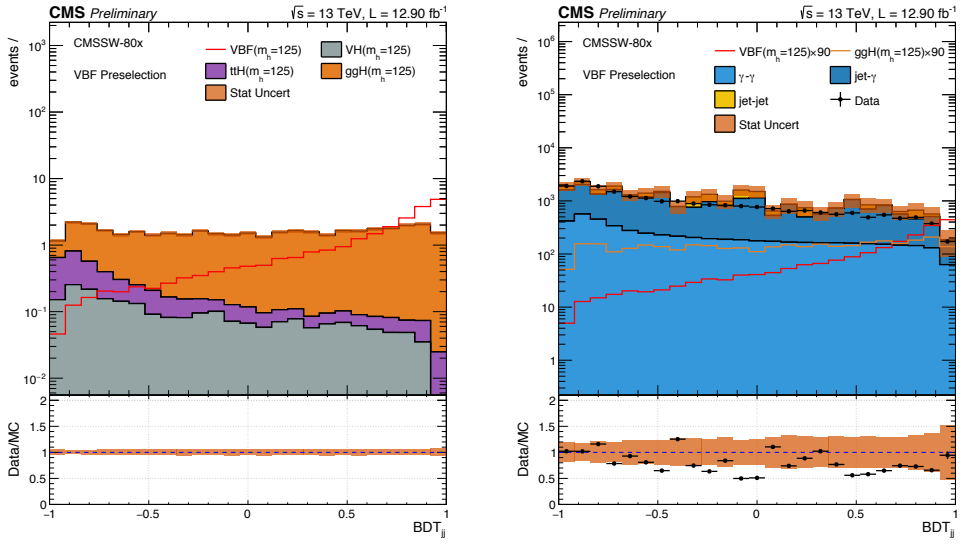
and are separated from PF photon candidates by $\Delta R > 0.5$. The leading (subleading) jet, i.e the one with the highest (second-highest) p_{T} in the event, is required to satisfy $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ ($p_{\mathrm{T}} > 20\,\mathrm{GeV}$). For events passing these requirements, an additional selection on the invariant mass of the *dijet* composed of the leading and subleading jets, m_{jj}, is imposed: $m_{\mathrm{jj}} > 250\,\mathrm{GeV}$.

For events passing the dijet preselection described above, the VBF-tagged categorisation proceeds as follows. First, a per-event BDT, which is hereafter referred to as BDT_{jj}, is used to identify VBF-like dijets. The BDT_{jj} does not have any knowledge of the quality and mass resolution of the diphoton, so a further per-event BDT, referred to as the $BDT_{\mathrm{jj},\gamma\gamma}$, is used to recover this information. The $BDT_{\mathrm{jj},\gamma\gamma}$ has the $BDT_{\gamma\gamma}$ and the BDT_{jj} output scores among its inputs variables. It is thus able to combine the VBF-like dijet identification power of one BDT with with mass resolution information from the other. A selection on the output score of the $BDT_{\mathrm{jj},\gamma\gamma}$ is then used to choose events for the VBF-tagged categories.

The BDT_{jj} is designed to use the kinematic properties of the diphoton-dijet system to identify VBF-like events where the Higgs boson decayed via $\mathrm{H} \to \gamma\gamma$. It is trained on simulated samples of diphoton events where the signal is defined as VBF ($\mathrm{H} \to \gamma\gamma$) events and the background consists of samples of SM events with a diphoton and a dijet in the final state, in addition to a simulated sample of ggH events where dijets are formed from pileup and initial or final state radiation. The input variables for this BDT are listed below:

- the invariant-mass-scaled transverse momentum ($p_{\mathrm{T}}/\,m_{\gamma\gamma}$) for the leading and subleading photons in the diphoton candidate;
- $p_{\mathrm{T}}^{\mathrm{j}_1}$ and $p_{\mathrm{T}}^{\mathrm{j}_2}$, the transverse momenta of the leading and subleading jets in the dijet;
- m_{jj};
- $|\eta_{\mathrm{j}_1} - \eta_{\mathrm{j}_2}|$, the separation of the jets in the dijet in the η-direction;
- $\eta^* = \eta_{\gamma_1+\gamma_2} - (\eta_{j_1} + \eta_{j_2})/2$, the *Zeppenfeld* variable [60] where $\eta_{\gamma_1+\gamma_2}$ refers to the η-position of the vector sum of the photon momenta;
- $|\phi_{\gamma\gamma} - \phi_{\mathrm{jj}}|$, the separation of the dijet and the diphoton in the ϕ-direction.

The distributions of the BDT_{jj} output scores for each simulated signal sample are shown in Figure 4.2a. The distributions of the BDT_{jj} output scores for data and simulated background samples (and some simulated signal samples) are shown in Figure 4.2b.



(a) BDT_{jj} output by production mode

(b) BDT_{jj} output comparing data and simulated signal and background

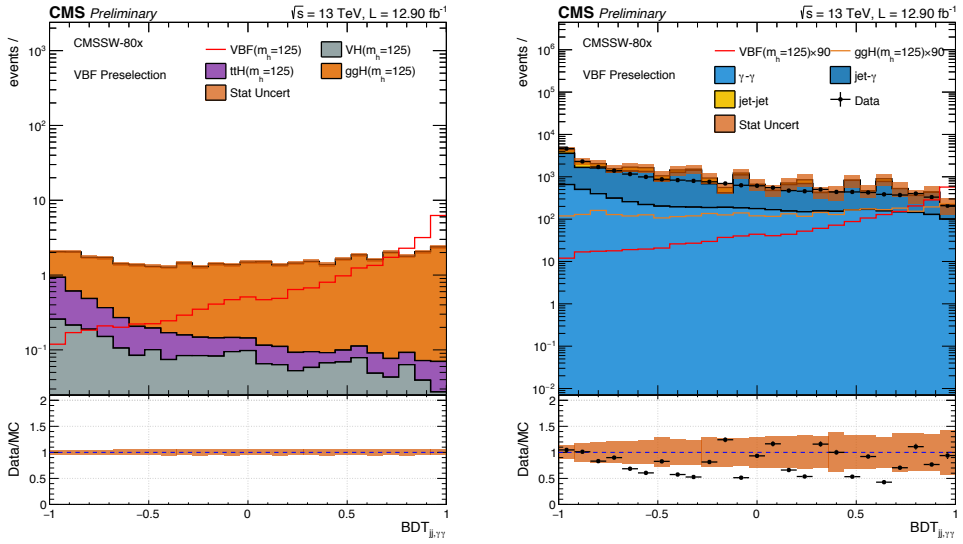
Figure 4.2: The output scores of the BDT_{jj} split by simulated production mode and comparing data and simulated signal and background.

The $BDT_{\mathrm{jj},\gamma\gamma}$ is designed to re-introduce information about the sensitivity of signal events while maintaining the power to identify VBF-like events. This is obtained by training the BDT on simulated events where the signal is a sample of VBF $\mathrm{H} \to \gamma\gamma$ events, while the background is composed of the SM diphoton background samples, as for the BDT_{jj} training. In this case, ggH events are treated as neither signal nor background. The inputs to the BDT are the following:

- the output score of the $BDT_{\gamma\gamma}$;
- the output score of the BDT_{jj};
- $p_{\mathrm{T}}^{\gamma\gamma} / m_{\gamma\gamma}$, the invariant-mass-scaled momentum of the diphoton system, which is included since it has a significant correlation to both the other inputs.

The distributions of the $BDT_{\mathrm{jj},\gamma\gamma}$ output scores for each simulated signal sample are shown in Figure 4.3a. The distributions of the BDT_{jj} output scores for data and simulated background samples (and some simulated signal samples) are shown in Figure 4.3b.

Selections on the $BDT_{\mathrm{jj},\gamma\gamma}$ output score are used to define two VBF-tagged categories. The most sensitive is referred to as the VBF-tagged 0 category and contains all the



(a) $BDT_{\mathrm{jj},\gamma\gamma}$ output by production mode **(b)** $BDT_{\mathrm{jj},\gamma\gamma}$ output comparing data and simulated signal and background

Figure 4.3: The output scores of the BDT_{jj} and $BDT_{\mathrm{jj},\gamma\gamma}$ split by simulated production mode and comparing data and simulated signal and background.

events with the highest $BDT_{\mathrm{jj},\gamma\gamma}$ output score above some boundary. The less-sensitive VBF-tagged 1 category contains all events for which the $BDT_{\mathrm{jj},\gamma\gamma}$ output score is below the first boundary, but still above a second boundary. Events for which the $BDT_{\mathrm{jj},\gamma\gamma}$ output score is below the second boundary fail the VBF-tagged categorisation, but may still be considered for inclusion in other analysis categories. The location of the two boundaries is optimised first by maximising the value of the signal-to-background ratio in the VBF-tagged 0 category, and then repeating the procedure after fixing the first boundary to maximise the signal-to-background ratio in the VBF-tagged 1 category. Of the simulated signal events which are categorised as VBF-like, the VBF-tagged 0 category contains approximately 72% VBF events and 27% ggH events, and the VBF-tagged 1 category contains approximately 55% VBF events and 43% ggH events.

4.4 VH-tagged categories

README this section will be updated with the final selections once the tags are finalised by Michael. This section uses the information from the legacy analysis as a placeholder.

Higgs boson events produced via the VH mode can be selected by identifying the decay products of the W or Z vector boson. There are three final states which can arise from these decays. First, the vector bosons can decay leptonically via $W \to \ell\nu$ or $Z \to \ell^+\ell^-$. Second, the vector bosons can decay to quarks which hadronize to form jets. Finally, the Z can decay to a pair of neutrinos, which are reconstructed as a large amount of missing energy. In this analysis, several categories are defined to target the final states listed above, labelled leptonic VH-tagged, hadronic VH-tagged and MET VH-tagged categories respectively. Although the addition of the VH-tagged categories does not significantly improve the overall significance of the analysis, it allows the measurement of the cross-section of the VH production mode. For all VH-tagged categories, events must mass the preselection defined in Section 3.3.3. To account for the fact that the p_{T} spectrum of Higgs decay particles is shifted towards higher values for VH events than for ggH events (due to recoil against the vector boson), the leading photon p_{T} requirement is increased to $p_{\mathrm{T}}/m_{\gamma\gamma} > 1/2$.

The leptonic VH-tagged categories target events with at least one charged lepton, and are further split into loose leptonic VH-tagged and tight leptonic VH-tagged categories in order to maximise the sensitivity. In both cases, fewer than three jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.4$ or within $\Delta R < 0.5$ of the photons or leptons must be reconstructed in the event, to avoid selecting events from the ttH process. The tight leptonic VH-tagged category targets the most sensitive events, i.e. those with a fully reconstructed final state (from Z decay) or large $\not{E}_{\mathrm{T}}$. Events which enter this category must satisfy the following requirements:

- the event must contain either: one reconstructed lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $\not{E}_{\mathrm{T}} > 35\,\mathrm{GeV}$, or two opposite-signed same-flavour leptons both with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and invariant mass in the range $[70, 110]\,\mathrm{GeV}$;
- the diphoton must pass a loose requirement on its $BDT_{\gamma\gamma}$ output score, which keeps X% of the signal while rejecting Y% of the background events.

Events falling in the loose leptonic VH-tagged category must satisfy a less stringent requirement on the amount of $\not{E}_{\mathrm{T}}$, but in this case additional selections must be added to suppress the SM background:

- the event must contain exactly one reconstructed lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and have $\not{E}_{\mathrm{T}} < 35\,\mathrm{GeV}$;
- the invariant mass of both possible electron-photon systems must be more than 10 GeV away from the mass of the Z boson. This is designed to reduce the contribu-

tion of Zγ events, where $\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$ and one of the electrons is misreconstructed as a photon;

- each photon in the diphoton must be separated from any reconstructed tracks by $\Delta R > 1.0$ to reduce the contribution from DY events where an electron is also misreconstructed as a photon;
- the diphoton must pass a loose requirement on its $BDT_{\gamma\gamma}$ output score, which keeps X% of the signal while rejecting Y% of the background events.

The MET VH-tagged category targets events where the Z decays to two neutrinos, and thus leaves no energy in the detector. This results in a large amount of reconstructed $\not{E}_{\mathrm{T}}$ in addition to the high-p_{T} photons from the Higgs boson decay. Events selected by the MET VH-tagged category must satisfy the following requirements, in addition to the same restriction on the number of jets as imposed on the leptonic VH-tagged categories:

- the event must contain $\not{E}_{\mathrm{T}} > 70\,\mathrm{GeV}$;
- the $\not{E}_{\mathrm{T}}$ direction and the momentum of the diphoton system must be separated in the ϕ-direction by at least 2.1, since they are expected to be back-to-back due to momentum conservation;
- the diphoton must pass a loose requirement on its $BDT_{\gamma\gamma}$ output score, which keeps X% of the signal while rejecting Y% of the background events.

Finally, the hadronic VH-tagged category targets events where the vector boson decays to quarks, leading to two jets in the event. The following selections are used for this category:

- the leading (subleading) photon must satisfy $p_{\mathrm{T}}/\,m_{\gamma\gamma} > 1/2$ ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$);
- the event must contain at least two jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$;
- the invariant mass of the dijet system must be in the range $[60, 120]\,\mathrm{GeV}$;
- the invariant-mass-scaled momentum of the diphoton system must be greater than $130\,m_{\gamma\gamma}/120$;
- the angle θ^{*} between the diphoton direction in the diphoton-dijet rest frame and the detector frame must satisfy $|\cos\theta^{*}| < 0.5$ to exploit differences in the angular correlation of the diphoton and the dijet for VH events.

Events which fail the selections for the VH-tagged categories may still be selected for other categories.

4.5 ttH-tagged categories

Higgs boson events are produced in association with a pair of top quarks in the ttH mode. The final state therefore contains two photons from the Higgs boson decay, as well as two b quarks and decay products from two W bosons. The W bosons will decay either hadronically (to quarks, which subsequently hadronize to form jets) or leptonically (to a lepton and corresponding neutrino). The distinctive topologies of these events can be used to identify Higgs boson events likely to have been produced by the ttH mechanism. The cross-section of ttH production is low, so the benefit to the analysis in terms of final significance of an observation is small. However, the categorisation of ttH-like events is important because it allows the measurement of the strength of the interaction of the Higgs boson with top quarks. Various extensions to the SM predict enhanced values of the strength of the ttH interaction, and such models can be tested through the experimental measurement of the cross-section of ttH events decaying to $\mathrm{H} \to \gamma\gamma$.

Two exclusive ttH-tagged categories are defined in this analysis. On the one hand, the leptonic ttH-tagged category aims to select ttH events where at least one of the W bosons decayed leptonically. On the other hand, the hadronic ttH-tagged category targets events where both W bosons decayed to quarks. In addition to usual preselection applied to candidate events, the requirement on the leading photon p_{T} is increased to $p_{\mathrm{T}} > m_{\gamma\gamma}/2$, for the same reason as described in Section 4.5.

The leptons which are used to choose events for the leptonic ttH-tagged category must satisfy certain requirements to be used for selection purposes. Muons are required to be within $|\eta| < 2.4$, pass a selection based on the properties of their tracks in the tracker and muon chambers and satisfy requirements on their pileup-corrected isolations. Electrons must pass loose identification requirements and the invariant mass of both possible electron-photon systems must be more than 10 GeV away from the mass of the Z boson. In order to be included in the leptonic ttH-tagged category, events must pass the following requirements:

- the diphoton must satisfy a loose selection on the $BDT_{\gamma\gamma}$ output which has approximately 70% signal selection efficiency and 15% background selection efficiency;
- the event must contain at least one selected lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$;
- the event must contain at least 2 jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $|\eta| < 2.4$ and separated by at least a distance $\Delta R = 0.4$ from a photon or lepton candidate;

- at least one of the jets should be tagged as a b jet using the CVSv2 algorithm medium requirement, as described in [61].

For events to be included in the hadronic ttH-tagged category, the following selections are made:

- the diphoton must satisfy a loose selection on the $BDT_{\gamma\gamma}$ output score which has approximately 95% signal selection efficiency and 45% background selection efficiency;
- there must be no leptons in the event which meet the requirements for the leptonic ttH-tagged category;
- there must be at least five jets in the event satisfying $p_{\mathrm{T}} > 25\,\mathrm{GeV}$;
- at least one of the jets should be tagged as a b jet using the CVSv2 algorithm medium requirement, as described in [61].

Events which fail the selections for the ttH-taggeds may still be selected for other categories.

4.6 Untagged categories

Diphotons which are not categorised into the VBF-tagged, VH-tagged or ttH-tagged categories can still be included in the so-called Untagged categories, which are split into subcategories to improve the overall sensitivity of the analysis. The splitting is performed by defining boundaries in the $BDT_{\gamma\gamma}$ output score distribution. The location of the boundaries as well as the number of subcategories is optimised using simulated signal and background samples which are independent from those used to train the $BDT_{\gamma\gamma}$. Since ggH events are produced with only a Higgs boson in the final state to first order, they are typically placed in the Untagged categories

The optimisation of the location of the boundaries is performed as follows. For a given number of subcategories N_{subcat}, the boundaries are initially spaced evenly throughout the $BDT_{\gamma\gamma}$ output score distribution. Events falling in the subcategory below the lowest boundary are discarded. Simplified models are used to parametrise the signal and background $m_{\gamma\gamma}$ distributions in the remaining categories. For the signal, the model is a sum of two Gaussian functions, which model the detector resolution and the uncertainty due to incorrect vertex assignment respectively. For the background, the $m_{\gamma\gamma}$ spectrum

is modelled using an exponential shape. The expected significance is then obtained by producing an Asimov dataset [62] from the signal and background models in each category, and then performing a simultaneous signal and background fit. The estimation of the expected significance is iteratively repeated, allowing the boundaries between the subcategories to float. The final set of boundaries is chosen such that the expected significance is maximised for a given N_{subcat}.

The procedure can be repeated separately for arbitrary values of N_{subcat}. In this analysis, $N_{\text{subcat}} = 4$ (ignoring the subcategory below the lowest boundary) was chosen, as moving to $N_{\text{subcat}} = 5$ produced a negligible improvement in the expected significance. The boundaries extracted from this optimisation procedure are represented by the vertical dashed lines in Figures 4.1a and 4.1b. The resulting Untagged categories are numbered from Untagged 0 (most sensitive) to Untagged 3 (least sensitive), and events where the diphoton has a $BDT_{\gamma\gamma}$ score too low to enter Untagged 3 are discarded.

4.7 Categorisation hierarchy

Each event can be assigned to only one category. To ensure this, a categorisation hierarchy is enforced. Each event is tested to see if it can be included in the first category in the hierarchy. If it satisfies the relevant requirements, the event enters the category and the categorisation is done. If not, the next category in the hierarchy is tested, and so on. If no further categories remain, the event is discarded. The hierarchy is as follows, with categories ordered from first tested to last tested: leptonic ttH-tagged, tight leptonic VH-tagged, loose leptonic VH-tagged, MET VH-tagged, hadronic ttH-tagged, hadronic VH-tagged, VBF-tagged 0, VBF-tagged 1, Untagged 0, Untagged 1, Untagged 2, Untagged 3.

Chapter 5

Statistical analysis

5.1 Signal modelling

Bibliography

[1] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[2] M. Thomson, "Modern particle physics". Cambridge University Press, New York, 2013.

[3] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[4] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9.

[5] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[7] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[8] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.

[9] E. Noether, "Invariant Variation Problems", *Gott. Nachr.* **1918** (1918) 235–257, doi:10.1080/00411457108231446, arXiv:physics/0503066. [Transp. Theory Statist. Phys.1,186(1971)].

[10] D. Griffiths, "Introduction to Elementary Particles". Physics textbook. Wiley, 2008.

[11] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (Jun, 1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[12] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[13] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[14] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[15] A. Salam, "Weak and electromagnetic interactions", in *Elementary particle physics: relativistic groups and analyticity*, N. Svartholm, ed., p. 367. Almqvist & Wiskell, Stockholm, 1968. Proceedings of the eighth Nobel symposium.

[16] U. M. Heller, M. Klomfass, H. Neuberger et al., "Numerical analysis of the Higgs mass triviality bound", *Nucl. Phys.* **B405** (1993) 555–573, doi:10.1016/0550-3213(93)90559-8, arXiv:hep-ph/9303215.

[17] S. L. Wu, "Brief history for the search and discovery of the Higgs particle A personal perspective", *Mod. Phys. Lett.* **A29** (2014), no. 09, 1330027, doi:10.1142/S0217732313300279, arXiv:1403.4425.

[18] OPAL Collaboration, DELPHI Collaboration, LEP Working Group for Higgs boson searches, ALEPH Collaboration, L3 Collaboration, "Search for the standard model Higgs boson at LEP", *Phys. Lett.* **B565** (2003) 61–75, doi:10.1016/S0370-2693(03)00614-2, arXiv:hep-ex/0306033.

[19] TEVNPH Collaboration, CDF Collaboration, D0 Collaboration, CDF Collaboration, "Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", 2012. arXiv:1203.3774.

[20] P. B. Renton, "Electroweak fits and constraints on the Higgs mass", in *Proceedings, 32nd International Conference on High Energy Physics (ICHEP 2004): Beijing, China, August 16-22, 2004*, pp. 564–567. 2004. arXiv:hep-ph/0410177.

[21] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[22] 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), no. 1, 1 – 29, doi:/10.1016/j.physletb.2012.08.020.

[23] ATLAS Collaboration and CMS 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.

[24] B. Mellado Garcia, P. Musella, M. Grazzini et al., "CERN Report 4: Part I Standard Model Predictions", Technical Report LHCHXSWG-DRAFT-INT-2016-008, CERN, Geneva, (May, 2016).

[25] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008), no. 08, S08001, doi:10.1088/1748-0221/3/08/S08001.

[26] LEP Injector Study Group, "LEP design report, volume I: The LEP injector chain; LEP design report, volume II: The LEP Main Ring". CERN, Geneva, 1983.

[27] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN.", Dec, 2008.

[28] M. Benedikt, P. Collier, V. Mertens et al., "LHC Design Report". https://cds.cern.ch/record/823808, 2004.

[29] CMS Collaboration, "CMS Luminosity - Public Results". http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[30] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

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

[32] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[33] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[34] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[35] CMS Collaboration, "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[36] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[37] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

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

[39] 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. [JINST8,9009(2013)].

[40] CMS Collaboration, "CMS ECAL first results with 2016 data". https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResultsCMSDPS2016031.

[41] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[42] CMS Collaboration, "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J.* **C60** (2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5,10.1140/epjc/s10052-009-1024-0. [Erratum: Eur. Phys. J. C61, (2009) 353].

[43] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[44] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[45] 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).

[46] 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).

[47] J. Alwall, R. Frederix, S. Frixione 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, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[48] 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.

[49] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[50] T. Gleisberg, S. Hoeche, F. Krauss et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[51] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[52] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.

[53] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[54] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, "The elements of statistical learning : data mining, inference, and prediction". Springer series in statistics. Springer, New York, 2009. Autres impressions : 2011 (corr.), 2013 (7e corr.).

[55] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.", *Ann. Statist.* **29** (10, 2001) 1189–1232, doi:10.1214/aos/1013203451.

[56] A. Hocker, J. Stelzer, F. Tegenfeldt et al., "TMVA - Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[57] R. Brun and F. Rademakers, "ROOT - An Object Oriented Data Analysis Framework", in *Proceedings AIHENP'96 Workshop, Lausanne*, volume 389 (1997) 81-86. September, 1996. See also http://root.cern.ch/.

[58] J. Gaiser, "Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'". PhD thesis, SLAC, 1982.

[59] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063.

[60] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color singlet exchange in Z + two jet events at the CERN LHC", *Phys. Rev.* **D54** (1996) 6680–6689, doi:10.1103/PhysRevD.54.6680, arXiv:hep-ph/9605444.

[61] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[62] G. Cowan, K. Cranmer, E. Gross et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J. C73, (2013) 2501].