Photon Identification Efficiency in $H \to \gamma\gamma$

Emily Orgill, The University of Manchester, United Kingdom

Supervised by Elisabeth Petit

September 11, 2014

Abstract

The photon identification efficiency in ATLAS for Higgs boson decays to diphotons is studied using Monte Carlo samples of all Higgs boson production modes at $\sqrt{s} = 8$ TeV. Isolated photon candidates are identified using a tight selection. Of particular interest is the effect of the jets produced with the Higgs boson on this identification. For leading photon transverse energies larger than 60 GeV, $\epsilon_{ID} > 94\%$. This Monte Carlo study suggests that the identification efficiency depends on the transverse energy of the photon, but not significantly dependent on the number of jets, the minimum separation between a photon and any jet, or in general on the production mode of the Higgs boson.

Contents

1 Introduction

A particle consistent with the Standard Model (SM) Higgs boson was reported in 2012 by both the ATLAS and CMS experiments [1, 2]. Since, further investigation has been undertaken to test the particle against SM Higgs predictions, such as couplings to fermions and vector bosons [3], and the spin [4]. None of the results to date show significant deviations from SM predictions.

This report focuses on Higgs boson decays to diphotons using Monte Carlo (MC) samples. In particular, it is concerned with the identification efficiency of photons. Photon identification is where events are selected such that they are very likely to be photons. It is important to perform this selection as photon signals can be faked by background jets. The small signal-to-background ratio of the Higgs boson at the LHC [5] necessitates accurate photon identification and jet rejection.

Other jets are also produced in many of the interactions whic also produce the Higgs boson (due to the parton nature of protons). Thus, an understanding of the jets and whether the jets influence the identification efficiency of photons is required, particularly if the photon and jet showers overlap. A pictorial representation of shower overlap, in η–ϕ space, is given in Figure 1. A correlation between the identification efficiency and the jets would be an undesirable effect as the jets are unpredictable.

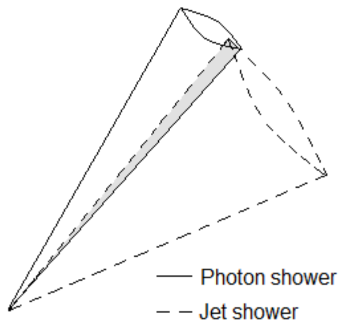


Figure 1: A jet and photon shower overlapping in the detector.

1.1 ATLAS detector

The ATLAS detector is a multi-purpose detector at the LHC in Geneva, Switzerland. The detector is forward-backward symmetric[1]; this makes it an ideal general purpose detector. Thus, the ATLAS experiment is involved in a wide range of searches and analyses, including that of the Higgs boson.

[1] ATLAS uses a right-handed coordinate system with its original at the nominal interaction point (IP) at the detector centre and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used, where ϕ is the azimuthal angle around the beam pipe. The pseudorapidity, η, is defined in terms of the polar angle, θ, as $\eta = -\ln \tan \theta/2$

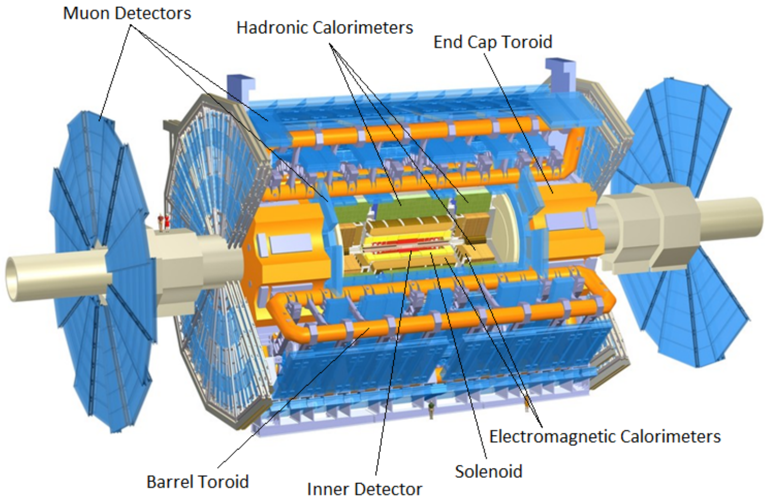


Figure 2: The ATLAS detector [6].

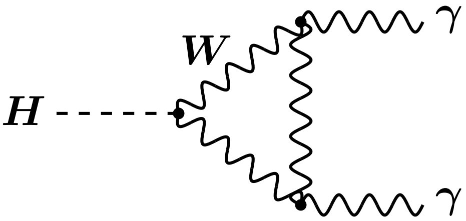
Figure 2 shows the main features of the detector. The inner detector provides precise tracking of particles close to the interaction point. The inner detector is surrounded by a solenoid which generates a 2 T magnetic field [7]. The liquid-argon (LAr) electromagnetic (EM) calorimeter covers the detector over the range $|\eta| < 3.2$, and is the part of the detector where the photon energy is deposited. It consists of a barrel, which covers $|\eta| < 1.475$, and two end-caps, covering $1.375 < |\eta| < 3.2$ [7]. The calorimeter has a fine granularity, which is good for precision measurements of photons. Precision is necessary for measuring the properties of the Higgs boson. Following the EM calorimeter is the hadronic calorimeter. This is where the majority of the jet energy is deposited. Finally, there are the toroid magnets, which provide the magnetic field to bend muons, which are detected in the muon detectors, which lie furthest from the detector.

1.2 Higgs boson

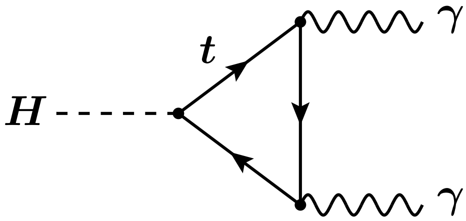
The Higgs boson is the particle associated with the Higgs field. When a particle interacts with this field, it acquires mass according to the strength of the interaction. Heavier particles have a stronger coupling to the Higgs field than lighter particles.

The Higgs boson has a mass of $m_H = 125.4$ GeV, as measured by ATLAS using $H \to \gamma\gamma$ and $H \to 4\ell$ [5]. This means that it can decay to all fundamental particles except the top quark, which has a much larger mass than that of the Higgs boson. The two most significant decay channels at the LHC are $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$. One reason for this is that there is a narrow mass peak in the invariant mass of the respective final states at the Higgs mass. There is also a relatively smooth background over which the Higgs mass can be extracted [5]. Henceforth, discussion will be limited to $H \to \gamma\gamma$.

The Higgs boson does not directly couple to photons as photons are massless. Instead, it decays to photons via a virtual W boson loop (Figure 3(a)) or via a virtual top quark loop (Figure 3(b)).



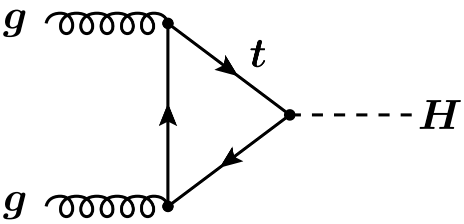
(a) $H \to \gamma\gamma$ via W loop



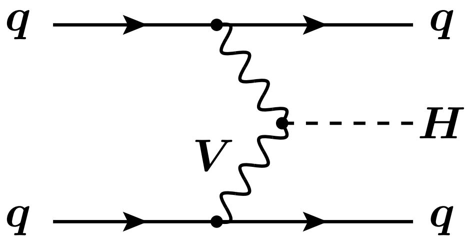
(b) $H \to \gamma\gamma$ via top quark loop

Figure 3: Feynman diagrams for Higgs production modes.

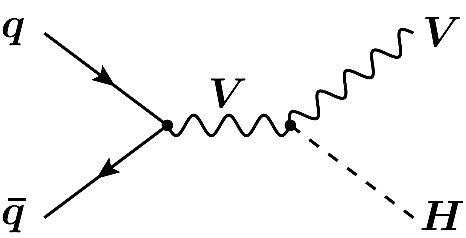
The Higgs boson has five production modes, which occur in different proportions of the total events. At the LHC, the dominant production mode is gluon-gluon fusion (ggF), as shown in Figure 4(a), via which 87.2% of all Higgs events are produced. The top quark, having the largest mass of any fundamental particle, couples most strongly to the Higgs, hence ggF, containing a virtual top-quark loop, is the most abundant. Weak vector boson fusion (VBF), shown in Figure 4(b), involves W or Z bosons and quarks in the intermediate state. W and Z bosons couple less strongly to the Higgs boson, and the initially interacting quarks must have sufficient energy to produce this intermediate state, and thus VBF occurs in 7.1% of all Higgs events.



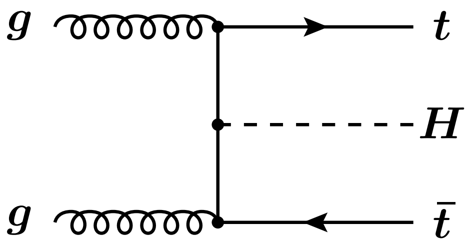
(a) Gluon-gluon fusion (ggF)



(b) Weak vector boson fusion (VBF)



(c) Higgstrahlung (WH/ZH)



(d) Associated production with $t\bar{t}$ (ttH)

Figure 4: Feynman diagrams for the Higgs boson production modes.

Higgstrahlung (WH/ZH) is shown in Figure 4(c). This mode is responsible for 5.1% of the Higgs bosons. It is less likely than VBF because there must be sufficient energy in the interaction for the virtual W or Z boson to decay into a W or Z, respectively, plus a Higgs boson. The lowest percentage, at 0.6%, is for Higgs bosons produced in association with $t\bar{t}$ (ttH), shown in Figure 4(d). In ttH, despite the strong coupling to the Higgs, the interacting gluons must have much higher energy originally in order to create the $t\bar{t}H$ intermediate state, and hence this production mode is very suppressed.

1.3 Photons

In $H \to \gamma\gamma$, the Higgs mass is determined by finding a resonance in the distribution of the invariant mass of the two-photon final state. This requires high precision as the signal is very small, and thus it is necessary to select events which have a high probability of being photons, and to reject as many background jet events as possible. This can be done by using the fact that photon and jet showers are different. Photons have a much narrower shower and almost all of the energy is deposited in the EM calorimeter. For jets, the showers are wider and the energy is largely deposited in the hadronic calorimeter. These differences can be used to determine candidate two-photon events.

1.3.1 Selection cuts

There are two different selections which are used to identify photons and reject fake signatures from jets. These are the loose and tight selections. Loose selection involves fewer discriminating variables than the tight selection. In the latter, more events which are possible fakes are cut (at the expense of losing some photons) and hence, there is a greater chance that the majority of events are photons. In this analysis, photons are identified using a tight selection cut. The variables used in this selection are shown in Table 1. These are based on the shower shapes in the calorimeter.

As well as the tight selection, for Higgs analysis there are further cuts imposed on the data. These are summarised in Table 2, where ΔR, the separation between photon and any jet in η–ϕ space, is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{1}$$

1.3.2 Photon identification efficiency

Using the variables described in Section 1.3.1, events are cut from the complete sample. The remaining events are considered most likely to be photons. Thus, the photon identification efficiency, ϵ_{ID}, is defined as:

$$\epsilon_{ID} = \frac{N^{tight,isolated}}{N^{total,isolated}} \tag{2}$$

where $N^{tight,isolated}$ is the number of isolated photons passing the tight selection, and $N^{total,isolated}$ is the total number of isolated photons.

Category	Description	Name
Acceptance	$\lvert\eta\rvert < 2.37$, with $1.37 < \lvert\eta\rvert < 1.52$ excluded.	-
Hadronic Leakage	Ratio of E_T in the first sampling of the hadronic calorimeter to E_T of the EM calorimeter (for $\lvert\eta\rvert < 0.8$ and $\lvert\eta\rvert > 1.37$).	R_{had_1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM calorimeter (for $0.8 < \lvert\eta\rvert < 1.37$).	R_{had}
EM Middle Layer	Ratio in η of cell energies in 3×7 and 7×7 cells	R_η
	Lateral width of the shower	ω_2
	Ratio in ϕ of cell energies in 3×3 and 3×7 cells	R_ϕ
EM Strip Layer	Shower width for the three strips around the maximum strip	ω_{s3}
	Total lateral shower width	ω_{stot}
	Fraction of energy within seven strips from the centre, not including the central three strips	F_{side}
	Energy difference between the energy in the second maximum in the strip layer, and the reconstructed energy in the strip with the minimal value between the first and second maxima	ΔE
	Ratio of energy difference between largest and next largest energy deposits	E_{ratio}

Table 1: Discriminating variables used in tight selection, from [8].

Description	Cut
Photon energy	$E_T > 25$ GeV
Photon isolation	$E_T^{iso} < 4$ GeV
Jet momentum	$p_T > 25$ GeV
Jet rapidity	$\lvert y\rvert < 4.4$ (equivalent to $\lvert\eta\rvert < 4.5$)
ΔR	$\Delta R < 0.4$

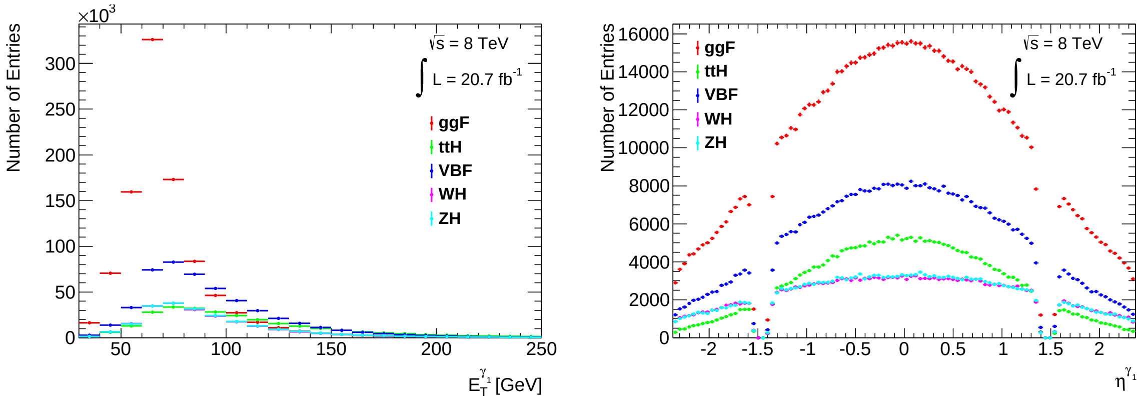
Table 2: Additional cuts on photons and jets required for Higgs analysis.

2 Results

2.1 Photon kinematics

The transverse energy of the leading photon, $E_T^{\gamma_1}$, and η^{γ_1} distributions are shown in Figures 5(a) and 5(b) respectively. The $E_T^{\gamma_1}$ peak in each production mode is in the range $60 < E_T^{\gamma_1} < 80$ GeV. As the leading photon carries away the largest fraction of the Higgs invariant mass, this is consistent with expectation.

Figure 5(b) shows that most photons are emitted at $\eta^{\gamma_1} = 0$, equivalent to emission perpendicular to the beam line. The distribution falls to zero at $|\eta| \simeq 1.5$, due to the



(a) Unnormalised distributions of $E_T^{\gamma_1}$.

(b) Unnormalised distributions of η^{γ_1}.

Figure 5: Plots of the leading photon kinematics for all production modes.

acceptance cut described in Section 1.3.1.

2.2 Identification efficiency vs photon transverse energy

First, ϵ_{ID} was plotted as a function of $E_T^{\gamma_1}$ for all production modes; the result of this is in Figure 6. Across all production modes, the efficiency increases rapidly below $E_T < 60$ GeV. Above this value, $\epsilon_{ID} > 94\%$. At large energies, ϵ_{ID} remains fairly constant, excepting statistical fluctuations. This trend is approximately the same across all production modes.

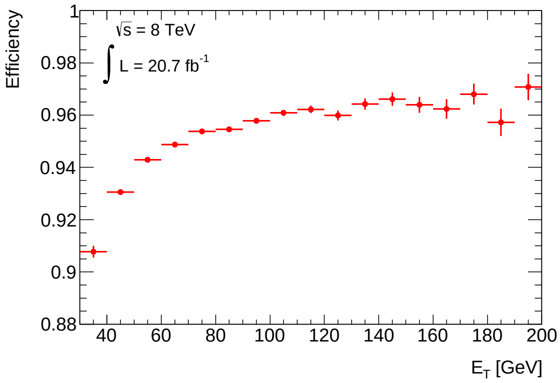
2.3 Jets

As mention in Section 1, the main focus of this analysis was to determine the effect, if any, of jets on ϵ_{ID}. If there is a significant reliance upon jets, this can cause problems with identifying photons well.

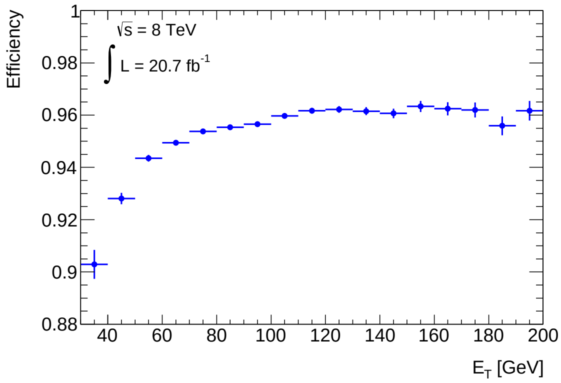
2.3.1 Jet distribution

Before ϵ_{ID} as a function of N_{jets} could be investigated, it was important to understand the N_{jets} spectrum for each production mode with reference to the Feynman diagrams for production in Figure 4.

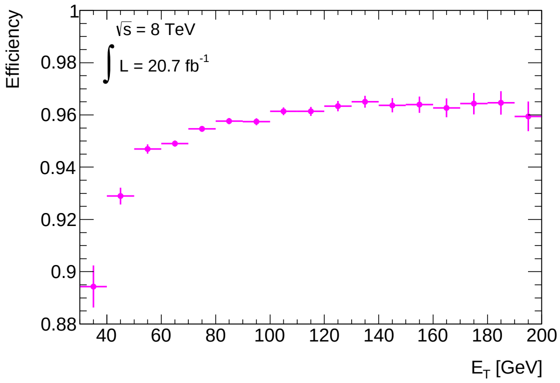
In ggF, the expected number of jets is zero at lowest order. This is verified by the results in Figure 7(a). However, the plot shows that $N_{jets} \neq 0$ in many events. There are two reasons for this. The first is due to higher orders, where gluons – which couple to both quarks and other gluons – are emitted and then decay to quarks, resulting in jets. The second is due to pile-up in the detector. Pile-up occurs as there are multiple interactions occurring per crossing. Any of these other interactions can produce jets, and because these interactions happen very close to the interaction of interest, it is difficult to



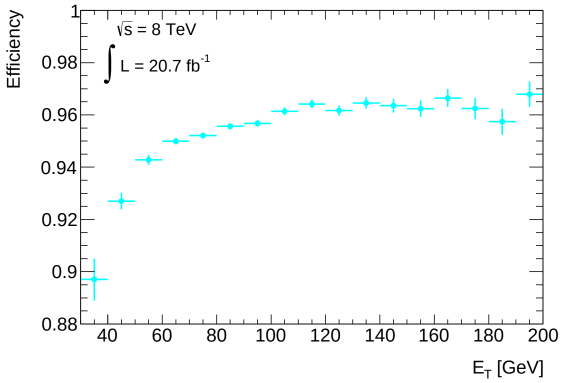
(a) ggF



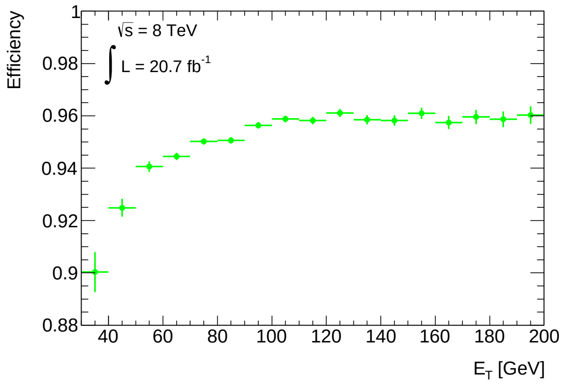
(b) VBF



(c) WH



(d) ZH

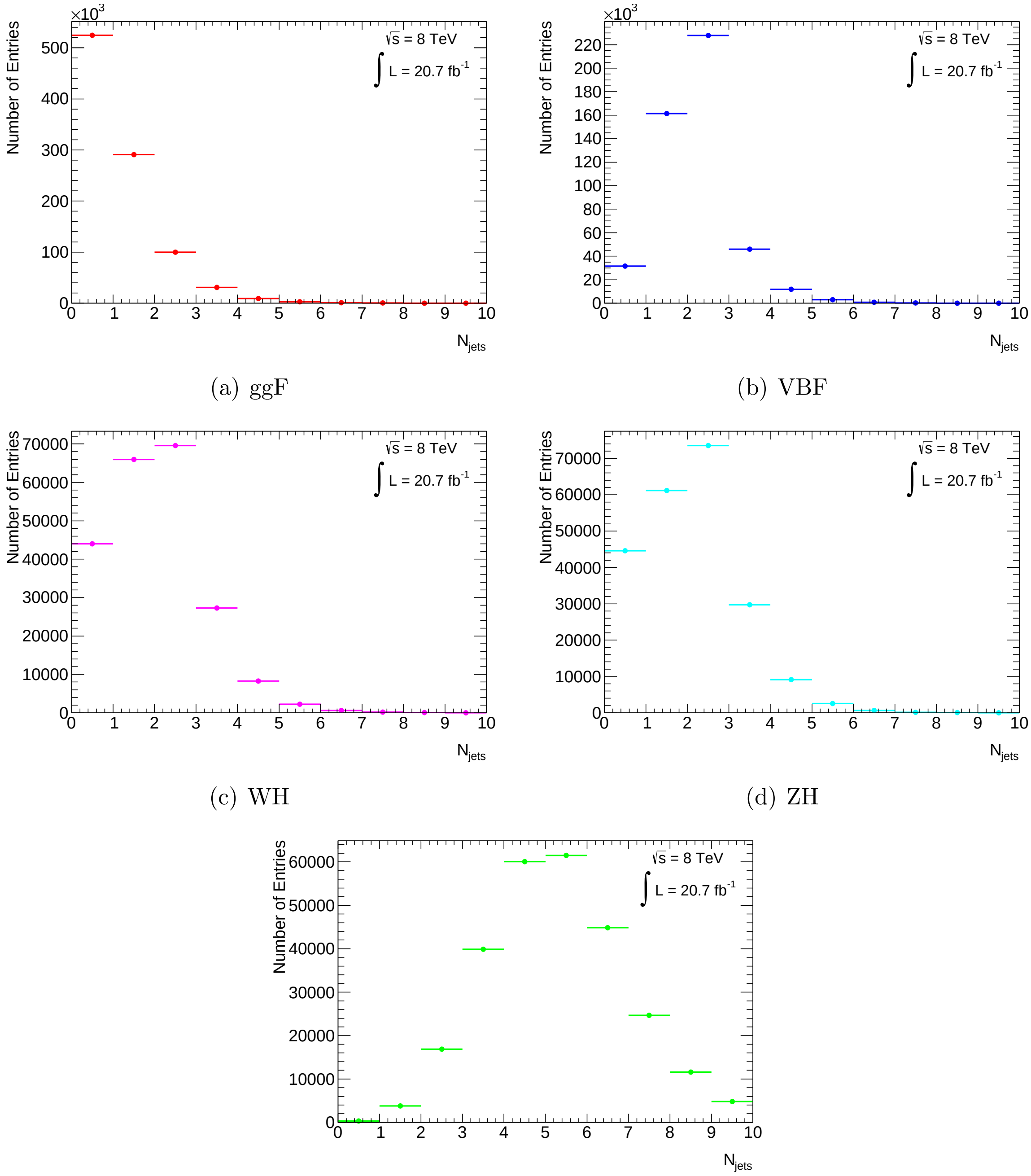


(e) ttH

Figure 6: ϵ_{ID} vs $E_T^{\gamma_1}$.

reconstruct these jets correctly to tracks originating from a different interaction. These two reasons for having more jets than expected are universal across all production modes.

As can be seen from the Feynman diagram for VBF in Figure 4(b), this production mode most commonly produces two jets. This expectation is verified by the MC results shown in Figure 7(b). This plot shows that there were also events with zero or one jets due to the jet cuts described in Table 2.



(a) ggF

(b) VBF

(c) WH

(d) ZH

(e) ttH

Figure 7: Distribution of the number of jets, N_{jets}.

WH and ZH, in Figures 7(c) and 7(d) respectively, have very similar distributions. The number of jets is dependent on the decay of the W or Z boson respectively [9]. In the majority of cases, both bosons decay to hadrons which results in two jets. Otherwise, for WH, ~30% decay to $\ell\nu$; for ZH, 10% decay to leptons and 20% to neutrinos. For

Mode	χ^2	N_{dof}	Slope	σ_{slope}	$\langle\epsilon_{ID}\rangle$	$\sigma_{\langle\epsilon_{ID}\rangle}$	Spread (%)
ggF	43.08	4	0.0042	0.0002	0.9490	0.0002	0.4583
VBF	11.29	4	0.0026	0.0003	0.9543	0.0003	0.2687
WH	7.01	4	0.0012	0.0004	0.9553	0.0004	0.1762
ZH	3.92	4	0.0018	0.0004	0.9545	0.0004	0.2213
ttH	11.00	8	0.0012	0.0002	0.9534	0.0004	0.2218

Table 3: Linear fit and mean statistics for ϵ_{ID} vs N_{jets} (based on Figure 8).

these, zero jets are produced.

The production mode in which the most jets are, on average, created is ttH. This can be seen from Figure 7(e). This is due to the top quarks, which do not hadronise and instead decay via $t \to bW$. The W boson decays as previously described for WH, but there are also two b-jets. Therefore 2, 4, and 6 jets will be produced with increasing probability. A possible reason why there are fewer 6-jet events than 4-jet events is the jet cuts, which also can explain why there are a lot of 5-jet events.

2.3.2 Identification efficiency vs number of jets

For each of the production modes, as shown in Figure 8, ϵ_{ID} appears to increase with N_{jets}. In order to check whether this increase is consistent with zero within the errors, a linear fit was performed.

The details of this linear fit are summarised in Table 3. ttH, WH and ZH have smaller slopes than ggF and VBF, thus there appears to be some dependence on production mode. In all cases, the slopes are not consistent with zero within the errors. The spread (in %) is defined as:

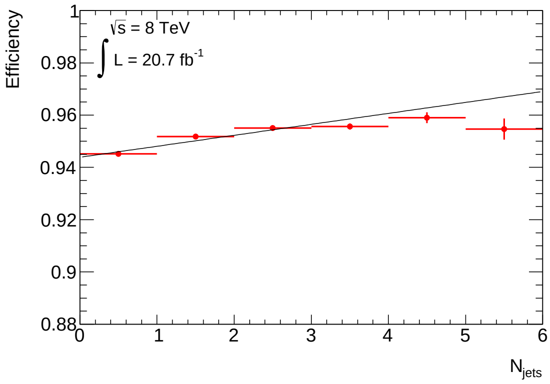
$$\text{spread (\%)} = 100 \times \frac{\sum_{i=0}^{N_{bins}} |x_i - \bar{x}|/\sigma_i}{\sum_{i=0}^{N_{bins}} 1/\sigma_i} \quad (3)$$

where N_{bins} is the number of bins, x_i is the efficiency in the ith bin, $\bar{x}$ is the mean value of the identification efficiency and σ_i is the statistical error in the ith bin. Table 3 shows that the spreads for the data are large, which is another indication that the change in ϵ_{ID} is significant. Thus, further investigation was required to see if this increase could be because of N_{jets} or because of other factors.

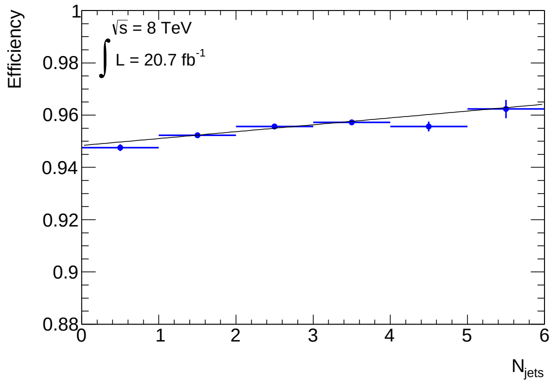
Based on the results in Section 2.2, the proposal was to find out if there was any dependence on $E_T^{\gamma_1}$ with N_{jets}. The results of this analysis are discussed in Section 2.3.3.

2.3.3 Mean photon transverse energy vs number of jets

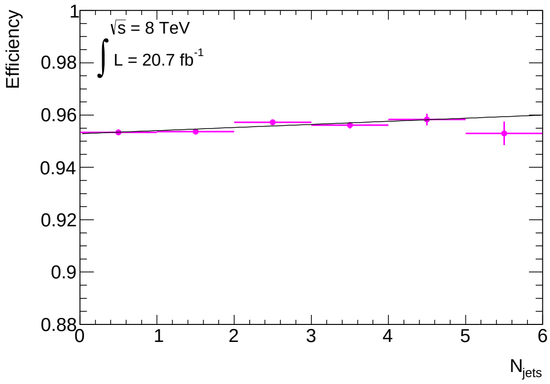
Figure 9 shows profile plots of the mean transverse energy of the photon, $\langle E_T^{\gamma_1}\rangle$, as a function of N_{jets}. In particular for ggF and VBF (Figures 9(a) and 9(b) respectively), the correlation between these quantities is strong. For the other production modes (Figures 9(c), 9(d) and 9(e)), the energy increase is much smaller. The increase in general is reasonable as, if there are more jets, the Higgs boson may be more boosted



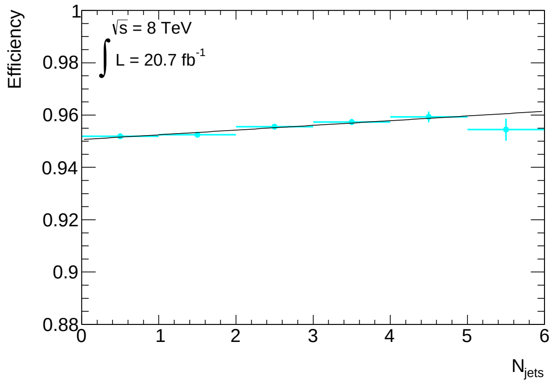
(a) ggF



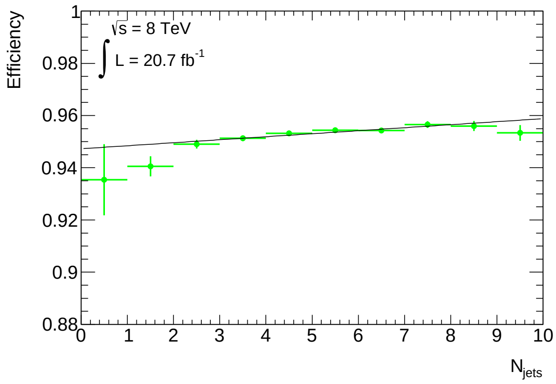
(b) VBF



(c) WH



(d) ZH

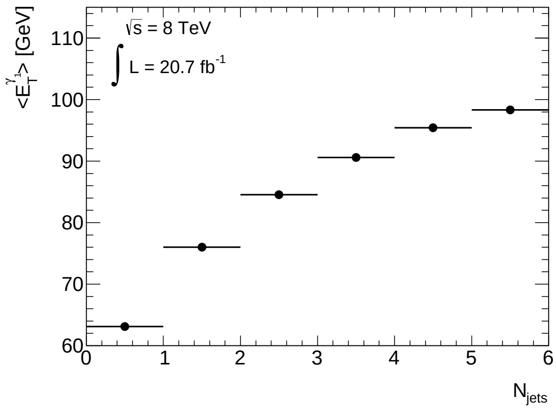


(e) ttH

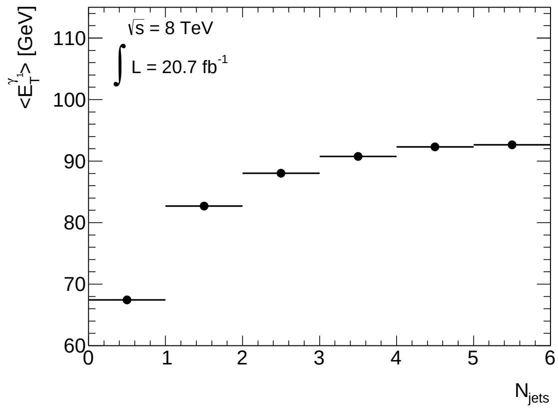
Figure 8: ϵ_{ID} vs N_{jets}.

in its rest frame ($p_T^H \neq 0$). Hence, the decay products (photons) would similarly be boosted. The differences in Figure 8 between production modes could therefore be due to the fact that the $\langle E_T^{\gamma_1} \rangle$ curves are different.

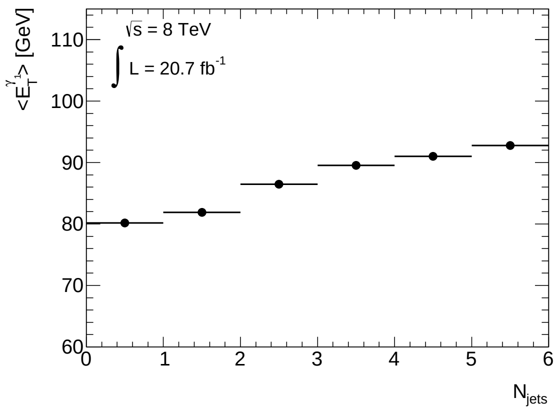
Since ϵ_{ID} increases more with N_{jets} for ggF and VBF, and less for the others, this is a good indication that the increase shown in Figure 8 is due to $E_T^{\gamma_1}$ and less likely to



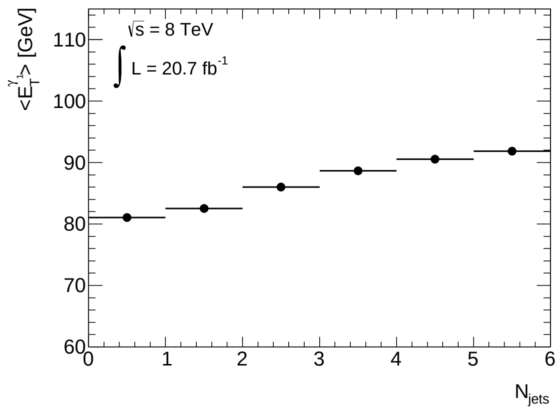
(a) ggF



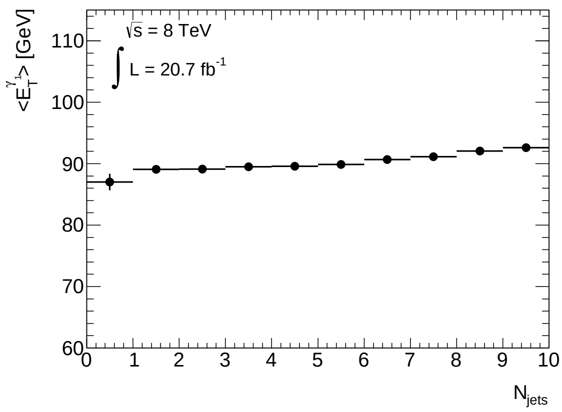
(b) VBF



(c) WH



(d) ZH



(e) ttH

Figure 9: $\langle E_T^{\gamma_1} \rangle$ vs N_{jets}.

be due to N_{jets} alone. However, N_{jets} is not the only feature of the jets which could affect the identification. Thus it is not possible yet to conclude that the identification is independent of the produced jets. In order to further examine the possible role of jets in the identification, it is necessary to consider ΔR from Equation 1 (see Section 2.4).

2.4 Minimum ΔR

ΔR_{min} is the minimum separation between the leading photon and the jets associated with the interaction. This is calculated by finding ΔR, using Equation 1, for the leading photon with respect to each jet and then rejecting all but the smallest value, which corresponds to the jet closest to the photon. This was calculated to check the likelihood of the photon and jet showers overlapping. If the showers overlap, this could affect the identification because the photon will not be as well isolated as assumed. The distribution for each production mode is shown in Figure 10.

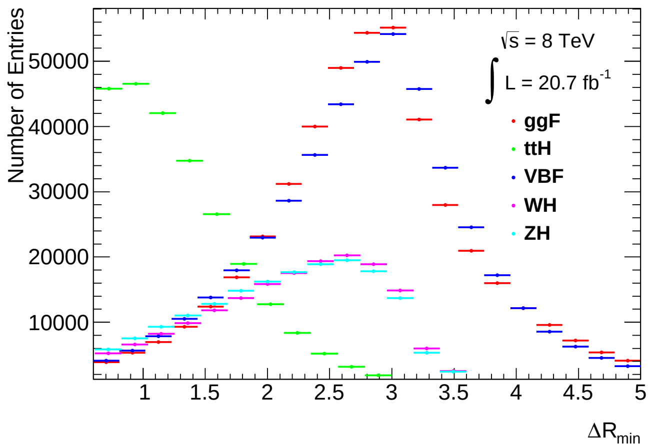


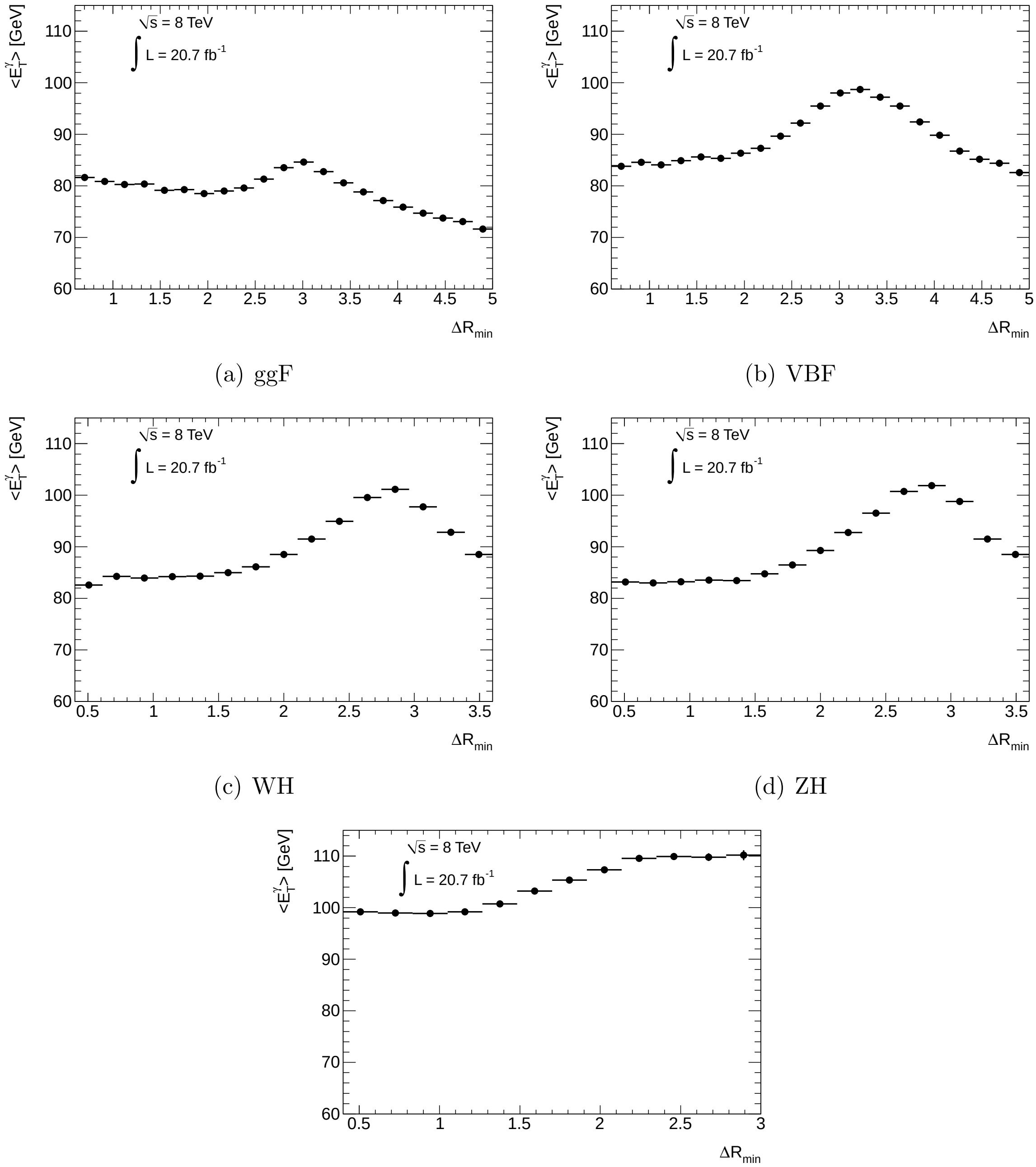
Figure 10: Distribution of ΔR_{min} for all production modes.

For all production modes except ttH, $\Delta R_{min} \simeq 3$ at the peak of the distribution. This is consistent with a back-to-back photon and jet. Thus, the expectation is that the jets in these productions should not significantly influence the identification because the showers will not usually overlap.

In ttH, however, the distribution peaks at $\Delta R_{min} \simeq 1$. Thus, the closest jet to the photon is not emitted back-to-back with the photon but is instead within a much smaller cone in η–ϕ space. Although events are removed if $\Delta R < 0.4$, where the jet and photon showers are likely to have a very large overlap, there is still a good chance that the showers have some overlap at $\Delta R \simeq 1$ because jet showers are wide. Therefore, if there is any correlation between ϵ_{ID} and ΔR_{min}, it is expected that it could be more pronounced for ttH.

2.4.1 Mean photon transverse energy vs minimum ΔR

First, $\langle E_T^{\gamma_1} \rangle$ was plotted as a function of ΔR_{min} to verify that the photon energy is not dependent on the separation. In each case, which can be seen in Figure 11, except in the region where the photons and jets are approximately back-to-back, the energy remains fairly constant. However, $\langle E_T^{\gamma_1} \rangle$ increases at larger distances in η–ϕ space, with a peak at $\Delta R_{min} \simeq 3$ in all cases including ttH.



(a) ggF

(b) VBF

(c) WH

(d) ZH

(e) ttH

Figure 11: $\langle E_T^{\gamma_1} \rangle$ vs ΔR_{min}.

The reason for why $\langle E_T^{\gamma_1} \rangle$ is not constant for all values of ΔR_{min} is unclear and requires further investigation. However, the fact that there is an increase only at large separation – where the jet and photon showers are very unlikely to overlap – is an indication that

Mode	χ^2	N_{dof}	Slope	σ_{slope}	$\langle\epsilon_{ID}\rangle$	$\sigma_{\langle\epsilon_{ID}\rangle}$	Spread (%)
ggF	38.03	19	0.00004	0.00039	0.9528	0.0003	0.1936
VBF	36.27	19	−0.0009	0.0004	0.9546	0.0003	0.1612
WH	8.05	13	0.0012	0.0007	0.9557	0.0005	0.1316
ZH	14.19	13	0.0004	0.0007	0.9552	0.0005	0.1533
ttH	4.86	10	0.0002	0.0008	0.9534	0.0004	0.0854

Table 4: Linear fit and mean statistics for ϵ_{ID} vs ΔR_{min} (based on Figure 12).

ϵ_{ID} does not significantly depend on ΔR_{min}. This is suggested even for ttH, despite its different ΔR_{min} distribution as described in Section 2.4.

2.4.2 Identification efficiency vs minimum ΔR

Figure 12 shows ϵ_{ID} as a function of ΔR_{min}. As with the dependence on N_{jets}, a linear fit analysis was undertaken. The results of this are summarised in Table 4.

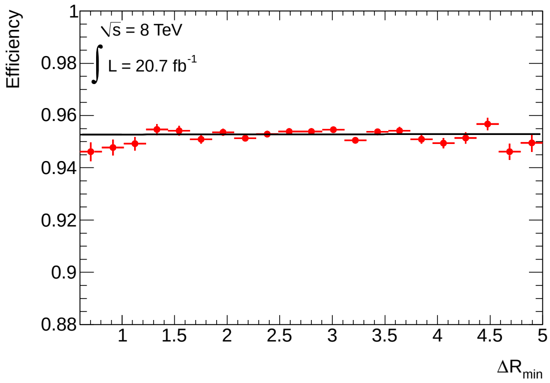
These show that the slopes are much flatter than in Section 2.3.2, as predicted. The slopes are also consistent with zero within the errors, and the spreads, calculated using Equation 3, are small. Thus it can be concluded that there is no dependence on ΔR_{min}. This is true even for ttH, so all production modes show a similar trend. This suggests that the isolation cut on photons is sufficient, and that, even if the showers could have some small overlap in ttH, there is no effect on the identification.

3 Conclusion

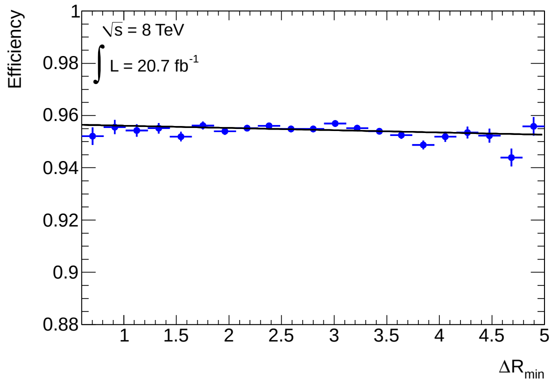
MC samples at $\sqrt{s} = 8$ TeV suggest that the photon identification in $H \to \gamma\gamma$ is dependent on $E_T^{\gamma_1}$, such that ϵ_{ID} increases rapidly with energy up to ~ 60 GeV and then becomes more constant, with $\epsilon_{ID} > 94\%$ for all production modes. The main focus of this analysis was to determine the effect of jets produced with the Higgs boson. The results suggest that there is no significant dependence on the number of jets nor the minimal separation between the photon and any jet. Hence, ϵ_{ID} is independent of a possible overlap between the jet and photon showers. It also suggests that $\langle E_T^{\gamma_1}\rangle$ is strongly dependent on the number of jets; there is also some change in $\langle E_T^{\gamma_1}\rangle$ for $\Delta R_{min} \simeq 3$ and the reason for this requires further investigation. These conclusions are the same for all production modes. However, as this is a MC study, the results discussed need to be compared to real data from ATLAS in order to test this conclusion.

Acknowledgements

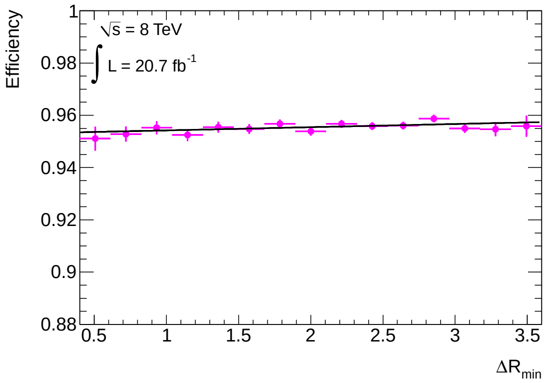
Many thanks to Elisabeth Petit, who provided me with an interesting project, patiently explain all aspects of the project, and was always ready to answer my questions. I learned a lot. Thanks also to the ATLAS group at DESY for being so welcoming and always willing to help me.



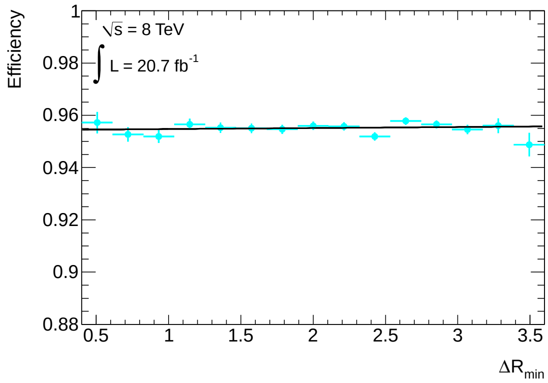
(a) ggF



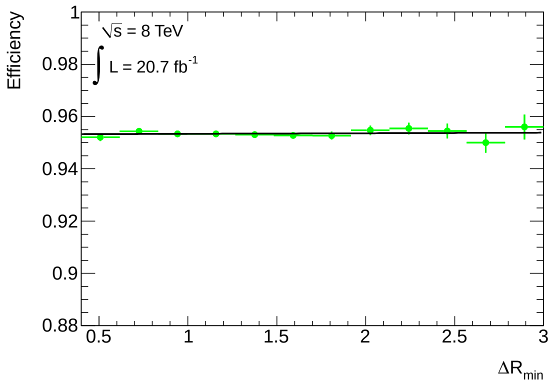
(b) VBF



(c) WH



(d) ZH



(e) ttH

Figure 12: ϵ_{ID} vs ΔR_{min}.

References

[1] The ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012), arXiv:1207.7214 [hep-ex].

[2] The CMS Collaboration, Phys. Lett. B **716**, 30 (2012), arXiv:1207.7235 [hep-ex].

[3] The ATLAS Collaboration, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, CERN-PH-EP-2014-198 (2014), submitted to Phys. Rev. D, arXiv:1408.7084 [hep-ex].

[4] The ATLAS Collaboration, Phys. Lett. B **726**, 120 (2013), arXiv:1307.1432 [hep-ex].

[5] The ATLAS Collaboration, *Measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data*, CERN-PH-EP-2014-122 (2014), submitted to Phys. Rev. D, arXiv:1406.3827 [hep-ex].

[6] The ATLAS Experiment, http://www.atlas.ch/photos/full-detector-cgi.html [Accessed 02/09/2014]

[7] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*. JINST **3** (2008) S08003.

[8] The ATLAS Collaboration, *Expected photon performance in the ATLAS experiment*, ATLAS NOTE.

[9] K.A. Olive *et al.* (Particle Data Group), Chin. Phys. C, **38**, 090001 (2014).