

# Properties of the Higgs-like state around 125 GeV in its decay into two photons at the CMS experiment

Matthew Kenzie

Imperial College London  
Department of High Energy Physics

CERN-THESIS-2014-090  
09/05/2014



A dissertation submitted to Imperial College London  
for the degree of Doctor of Philosophy



## Abstract

Results are presented of a search for the Standard Model (SM) Higgs boson decaying into two photons at the Compact Muon Solenoid (CMS) experiment housed at the Large Hadron Collider (LHC), CERN. An excess of events is observed over the background expectation with a local significance of  $5.7\sigma$ , where the SM expectation is  $5.2\sigma$ , constituting a standalone discovery of the particle first observed by the ATLAS and CMS experiments in July 2012. Measurements of the particle's signal strength, mass and couplings are presented along with an analysis of its spin. The results show a high level of compatibility with the predictions for a SM Higgs boson. The observed state's signal strength relative to the SM expectation is found to be  $\sigma/\sigma_{\text{SM}} = 1.14^{+0.26}_{-0.23}$ . The observed state's mass is found to be  $124.72 \pm 0.35$  GeV. The signal strength relative to the SM expectation when probing production mechanisms through fermionic modes only is  $1.13^{+0.37}_{-0.31}$ , and from bosonic production modes only is  $1.16^{+0.63}_{-0.57}$ . A spin-2 graviton, produced entirely by gluon fusion, is excluded at 94% C.L. (92% expected) and a spin-2 graviton, produced entirely by quark-antiquark annihilation, is excluded at 85% C.L. (83% expected).





## Declaration

This dissertation is not the result of entirely my own work. The concepts and ideas described in Chapters 2 and 3, whilst my own words, are based on the work of others. Considerable parts of Chapters 4 and 5, specifically Sections 4.3, 4.4, 4.5 and 5.1, are produced in collaboration with other members of the CMS  $H \rightarrow \gamma\gamma$  group and wider CMS collaboration. Consequently some of the techniques and ideas presented are not entirely my own. The majority of everything presented in Chapters 6 and 7 is my own work. Where appropriate ideas from others are referenced and figures from other sources are labelled with “ $H \rightarrow \gamma\gamma$ ”, “LHC” or “CMS” and the source is referenced in the figure captions.

This dissertation has not been submitted for another qualification to this or any other university and does not exceed the word limit for the respective Degree Committee.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Matthew Kenzie



## Acknowledgements

Firstly I would like to thank my supervisor Paul Dauncey, mainly for the numerous meals, coffees and beers he has bought me but also for his exceptional academic mentoring, shared knowledge and personal support. Cheers for putting up with my laziness especially in administrative matters and paperwork. Thanks also to my other (unofficial) supervisor Chris Seez for his advice, both scientific and political, and for many stimulating discussions, even though I have still not been invited to share a cigar! I owe a huge amount to Nick Wardle for his close collaboration and help over the years. I would also like to thank the CMS  $H \rightarrow \gamma\gamma$  group, some of whose hard fought analysis techniques and figures are used in this thesis. The many interesting discussions (arguments) and ideas have dictated my enjoyment of particle physics and desire to stay in the field. Thanks to the Science and Technology Facilities Council (STFC) and the Grundy Educational Trust for funding my PhD.

Thank you to all of my friends, who have endured my dreadful sense of humour and general lack of respect for punctuality, tidiness and organisation. I can't "name check" everyone but to Darren, Patrick and Andrew - thanks for your time. To Emma, thank you so much for putting up with me, especially over long distance, and thanks for allowing me to follow my passion whilst still maintaining a (semi) functional relationship with you. Lastly, to my family. My brothers have helped me see the amusing side of "a post-doc" and without the support of my parents I would simply not have had the opportunity to do this. For that, and all the other countless things, I will be eternally grateful.



*“For my father”*

—



*“Data! data! data! I cannot make bricks without clay”*

— The Adventures of Sherlock Holmes, Sir Arthur Conan Doyle





# Contents

List of figures	xvii
List of tables	xxi
List of Acronyms	xxii
<b>1. Introduction</b>	<b>3</b>
<b>2. Theory</b>	<b>5</b>
2.1. The Standard Model . . . . .	5
2.1.1. Fundamental particles and forces . . . . .	5
2.1.2. Gauge Theories . . . . .	7
2.1.3. Quark and Lepton states . . . . .	9
2.1.4. Electroweak symmetry breaking . . . . .	10
2.2. Higgs production at the LHC . . . . .	13
2.3. Higgs decay into two photons . . . . .	15
2.3.1. Backgrounds to the $H \rightarrow \gamma\gamma$ decay at the LHC . . . . .	17
<b>3. The CMS experiment</b>	<b>19</b>
3.1. The LHC . . . . .	19
3.2. The CMS detector . . . . .	20
3.2.1. Tracking system . . . . .	22
3.2.2. Electromagnetic calorimeter . . . . .	24
3.2.3. Hadronic calorimeter . . . . .	29
3.2.4. Muon Chambers . . . . .	29
3.3. Photon reconstruction . . . . .	30
3.3.1. Electron and photon differences . . . . .	33
3.4. Particle flow and jets . . . . .	33

3.5.	Isolation . . . . .	35
3.6.	Pileup . . . . .	36
<b>4.</b>	<b>Common analysis components</b>	<b>37</b>
4.1.	Boosted Decision Trees . . . . .	38
4.1.1.	Classification BDT . . . . .	38
4.1.2.	Regression BDT . . . . .	40
4.2.	Data samples and triggering . . . . .	40
4.2.1.	Monte Carlo Simulation . . . . .	41
4.2.2.	Pileup and beamspot reweighting . . . . .	42
4.3.	Energy measurement of photons . . . . .	44
4.3.1.	Correcting for residual discrepancies between data and Monte Carlo simulation . . . . .	46
4.4.	Vertex reconstruction . . . . .	48
4.4.1.	Estimating the per-event probability that the correct vertex is chosen	52
4.5.	Event preselection . . . . .	54
4.6.	Using $Z$ decays for validation and efficiency measurements . . . . .	55
<b>5.</b>	<b>Selection and Categorisation</b>	<b>57</b>
5.1.	Event selection . . . . .	58
5.1.1.	Selection using cuts in categories . . . . .	58
5.1.2.	Photon ID MVA . . . . .	59
5.1.3.	Diphoton event-level MVA . . . . .	62
5.2.	Event categorisation . . . . .	67
5.2.1.	Exclusive mode tagging . . . . .	69
5.2.2.	Inclusive mode categorisation and VBF dijet categorisation in the mass-factorised MVA analysis . . . . .	73
5.2.3.	Inclusive mode categorisation in the sideband MVA analysis . . . .	74
5.2.4.	Inclusive mode categorisation in the spin analysis . . . . .	75
5.2.5.	Event categorisation summary . . . . .	80
<b>6.</b>	<b>Analysis</b>	<b>83</b>
6.1.	Signal modelling . . . . .	83
6.1.1.	Mass-factorised analysis . . . . .	84
6.1.2.	Sideband analysis . . . . .	89
6.1.3.	Spin analysis . . . . .	89

6.2. Background modelling . . . . .	90
6.2.1. Mass-factorised analysis . . . . .	90
6.2.2. Sideband analysis . . . . .	98
6.2.3. Spin analysis . . . . .	102
6.3. Systematic Uncertainties . . . . .	103
6.3.1. Implementation of the systematic uncertainties . . . . .	103
6.3.2. Systematic uncertainties related to individual photons . . . . .	106
6.3.3. Systematic uncertainties related to diphoton events . . . . .	108
6.3.4. Systematic uncertainties related to production mode tagged classes	110
6.3.5. Summary . . . . .	110
6.4. Statistical interpretation of the data . . . . .	111
<b>7. Results</b>	<b>117</b>
7.1. Best fit model to data . . . . .	118
7.2. Exclusion limits and $p$ -value . . . . .	118
7.3. Measurement of physical parameters . . . . .	127
7.4. Spin . . . . .	130
7.4.1. SM compatibility check . . . . .	130
7.4.2. Hypothesis tests of the SM Higgs, $0^+$ , vs. graviton-like, $2_m^+$ . . . .	132
<b>8. Conclusions</b>	<b>137</b>
<b>Bibliography</b>	<b>139</b>
<b>A. Envelope function choices</b>	<b>145</b>



# List of figures

2.1. Feynman diagrams for SM Higgs production at the LHC . . . . .	14
2.2. Generator level Higgs distributions . . . . .	15
2.3. SM Higgs production cross section at the LHC . . . . .	16
2.4. Feynman diagrams for the SM Higgs to two photon decay at leading order.	16
2.5. SM Higgs branching fraction . . . . .	17
2.6. Higgs to two photon backgrounds at the LHC . . . . .	18
3.1. CMS diagram . . . . .	20
3.2. CMS tracker . . . . .	23
3.3. Vertex resolution . . . . .	23
3.4. CMS tracker material budget . . . . .	24
3.5. CMS ECAL schematic . . . . .	25
3.6. ECAL energy resolution . . . . .	27
3.7. ECAL laser corrections . . . . .	28
3.8. Particle flow jet resolution . . . . .	30
3.9. A schematic of the domino clustering algorithm . . . . .	32
4.1. The total integrated luminosity delivered to and recorded by CMS during the 2011 and 2012 run periods . . . . .	41
4.2. Distribution of the number of reconstructed vertices . . . . .	43

4.3. Distribution of $\Delta z$ (the distance between the chosen vertex and the true vertex in the $z$ direction) . . . . .	43
4.4. A comparison of the predicted probability density of $E_{true}/E_{raw}$ from the regression training to the distribution in a statistically independent sample	46
4.5. The $Z \rightarrow e^+e^-$ invariant mass shape before and after scale and smearing corrections are applied . . . . .	48
4.6. The $Z \rightarrow e^+e^-$ invariant mass distribution at 8 TeV when the electrons are reconstructed as photons . . . . .	49
4.7. A schematic showing the two methods used for locating the primary vertex	50
4.8. The vertex BDT response for $Z \rightarrow \mu^+\mu^-$ events . . . . .	52
4.9. The efficiency of locating the correct vertex . . . . .	53
4.10. A comparison of the true vertex efficiency and the average vertex probability	54
5.1. Cut based photon ID efficiency as measured in $Z \rightarrow e^+e^-$ MC simulation	60
5.2. The output distribution of the photon identification BDT . . . . .	63
5.3. The output distribution of the photon identification BDT in $Z \rightarrow e^+e^-$ decays . . . . .	63
5.4. The diphoton BDT response . . . . .	66
5.5. The diphoton BDT response in $Z \rightarrow e^+e^-$ decays . . . . .	68
5.6. The kinematic dijet BDT response . . . . .	71
5.7. The distribution of the combined dijet-diphoton BDT response . . . . .	71
5.8. Two dimensional distributions of the diphoton BDT output and $\Delta m/m_H$ for the sideband analysis . . . . .	76
5.9. The inclusive category bin definitions for the SMVA analysis . . . . .	77
5.10. The distribution of $ \cos(\theta_{CS}^*) $ before and after selection cuts for different spin signals . . . . .	78
5.11. Acceptance $\times$ efficiency ratio between the $2_m^+$ (gluon fusion production) and $0^+$ (all SM production modes) of the spin analysis event selection . .	79

6.1.	Diphoton invariant mass distributions for the datasets at 7 and 8 TeV . .	84
6.2.	The efficiency $\times$ acceptance of the analysis selection for SM Higgs MC . .	85
6.3.	The diphoton invariant mass shape for SM Higgs signal . . . . .	86
6.4.	The composition and width of the signal in each analysis category . . . .	88
6.5.	Conceptual idea of the method of likelihood profiling . . . . .	92
6.6.	A toy example of using the envelope method for estimating the background	93
6.7.	An estimation of the correction required in the envelope method . . . . .	95
6.8.	The invariant mass distribution for the 8 TeV dataset . . . . .	99
6.9.	An example of one of the sideband fits for the SMVA . . . . .	100
6.10.	The distribution of data, background and signal for the SMVA . . . . .	101
6.11.	A demonstration of the test statistic distribution . . . . .	113
7.1.	The diphoton invariant mass distribution in data with the best fit signal plus background overlaid for the untagged and dijet tagged categories in the 8 TeV dataset . . . . .	119
7.2.	The diphoton invariant mass distribution in data with the best fit signal plus background overlaid for the $VH$ and $t\bar{t}H$ -tagged categories in the 8 TeV dataset . . . . .	120
7.3.	The diphoton invariant mass distribution in data with the best fit signal plus background overlaid for the untagged and dijet 0 categories in the 8 TeV dataset . . . . .	121
7.4.	The diphoton invariant mass distribution in data with the best fit signal plus background overlaid for the dijet 1,2 and $VH$ -tagged categories in the 8 TeV dataset . . . . .	122
7.5.	The diphoton invariant mass distribution in data with the best fit signal plus background overlaid for the $t\bar{t}H$ -tagged categories in the 8 TeV dataset	123
7.6.	The unweighted diphoton invariant mass distribution for all analysis events in data with the combined best fit signal plus background model overlaid	123

7.7. The $S/(S+B)$ weighted diphoton invariant mass distribution for all analysis events in data with the combined best fit signal plus background model overlaid . . . . .	124
7.8. The expected and observed exclusion limits for a SM Higgs boson at 95% confidence level . . . . .	125
7.9. The expected and observed local $p$ -value to reject the background hypothesis	126
7.10. The observed best fit value of the signal strength modifier $\mu$ . . . . .	127
7.11. The 1D $-2\Delta LL$ scan of the signal strength modifier $\mu$ . . . . .	128
7.12. 2D $-2\Delta LL$ scans of $\mu$ and $m_H$ . . . . .	129
7.13. Best fit values and $-2\Delta LL$ scans of the signal strength from fermionic and bosonic production modes . . . . .	131
7.14. The observed compatibility of the signal strength between channels . . .	132
7.15. The SM signal strength extraction in bins of $ \cos(\theta_{CS}^*) $ for the spin analysis	133
7.16. Distributions of the test statistic for different spin hypotheses compared to the SM . . . . .	134
7.17. Distribution of the test statistic as a function of $f_{q\bar{q}}$ for the spin analysis	135
A.1. The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the inclusive, dijet and vector boson associated production ( $VH$ ) lepton tag categories in the 7 TeV dataset. . . . .	146
A.2. The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the $VH \cancel{E}_T$ and jet tag and top quark associated production ( $t\bar{t}H$ ) categories in the 7 TeV dataset.	147
A.3. The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the inclusive and vector boson fusion ( $VBF$ ) dijet tag categories in the 8 TeV dataset. . . . .	148
A.4. The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the $VH$ and $t\bar{t}H$ tag categories in the 8 TeV dataset. . . . .	149



# List of tables

2.1. The fundamental matter particles . . . . .	6
2.2. A summary of the fundamental force-carrying particles in the SM . . . .	7
4.1. Preselection cut values. . . . .	56
5.1. Photon ID selection cut values. The cuts are applied to both the leading and subleading photons. . . . .	60
5.2. Definition of the photon resolution categories in the spin analysis . . . .	79
5.3. Definition of diphoton $ \cos(\theta_{\text{CS}}^*) $ categories in the spin analysis . . . . .	80
5.4. The event classes at 7 and 8 TeV and some of their main selection require- ments . . . . .	81
6.1. Expected number of SM Higgs boson events at $m_H = 125$ GeV and estimated background density at $m_{\gamma\gamma} = 125$ GeV . . . . .	87
6.2. A summary of all the systematic uncertainties considered in the analysis.	115
6.3. Magnitude of the uncertainty on the signal strength, $\sigma/\sigma_{\text{SM}}$ , induced by the systematic uncertainties on the signal . . . . .	116
6.4. Magnitude of the uncertainty on the signal position, $m_H$ , induced by the systematic uncertainties on the signal . . . . .	116
7.1. The values of the best fit signal strength, $\sigma/\sigma_{\text{SM}}$ , when $m_H$ is profiled, for the 7 TeV, 8 TeV and combined datasets . . . . .	128



# List of Acronyms

APD	- avalanche photodiode
BDT	- Boosted Decision Tree
CiC	- Cuts in Categories
CMS	- Compact Muon Solenoid
DT	- Decision Tree
ECAL	- electromagnetic calorimeter
GED	- global event description
$ggH$	- gluon fusion
HCAL	- hadronic calorimeter
LHC	- Large Hadron Collider
MVA	- Multivariate Analysis
MC	- Monte Carlo
MFM	- Mass-Factorised MVA
$\text{PbWO}_4$	- lead tungstate
PF	- particle flow
POI	- parameter of interest
QCD	- quantum chromodynamics
QFT	- Quantum Field Theory

SM	- Standard Model
SMVA	- Sideband MVA
$t\bar{t}H$	- top quark associated production
$VBF$	- vector boson fusion
VEV	- vacuum expectation value
VPT	- vacuum phototriodes
$VH$	- vector boson associated production
$WH$	- $W^\pm$ boson associated production
$ZH$	- $Z$ boson associated production





# Chapter 1

## Introduction

The discovery of a new particle with a mass around 125 GeV in the search for the SM Higgs boson was announced simultaneously by the ATLAS and CMS collaborations in July 2012[1, 2]. Since then, about three times more data has been taken at the LHC before the running period including 2011 and 2012 (referred to as “Run 1”) terminated in December 2012 for scheduled maintenance and upgrades. One of the primary goals of the Higgs groups at the LHC is now to study the properties of this new resonance and determine whether it is the SM Higgs boson or not. Some of the work in this thesis contributed towards the discovery in 2012 but much of the work detailed here builds upon that and focuses on measuring the properties of the new state in its decay into two photons.

Chapter 2 gives a description of the fundamental constituents of matter and the forces that govern their interactions. The framework which underlies our theoretical predictions is known as the SM of particle physics and the concepts behind the theory, in the context of local gauge symmetries, are explained. The mechanism by which the fundamental particles acquire a mass, spontaneous symmetry breaking (or alternatively the Higgs mechanism), is summarised and serves as a motivation for the existence of, and consequent desire to search for, the Higgs boson. The chapter concludes by discussing Higgs boson production at the LHC and its decay into two photons with a focus on the predominant backgrounds for this search and how one can measure its spin.

In Chapter 3 the main apparatus for the analysis, the CMS detector, is detailed. There is an explanation of the main detector subsystems with a particular focus on the electromagnetic calorimeter (ECAL) which is used to measure photon energies. A short

description of some of the physics object reconstruction essentials is also given; particle flow, jets, isolation and pileup.

This thesis presents two complementary analyses and an additional analysis tailored to separate between different Higgs spin hypotheses. Chapter 4 gives a description of the common analysis elements which are shared by all three. The topics covered include the datasets and Monte Carlo (MC) simulation, photon energy measurement and primary vertex location of the Higgs decay. There are also some other useful preliminary topics discussed; the use of Boosted Decision Trees (BDT) and the use of the  $Z \rightarrow e^+e^-$  decay as a control source for the  $H \rightarrow \gamma\gamma$  analyses.

Chapter 5 gives a description of the event selection used in the three analyses and explains how the events are split into categories in order to improve sensitivity and help to reduce the errors on Higgs couplings measurements.

The full details of the statistical treatment of the data are explained in Chapter 6, which also includes a description of the signal and background modelling and the treatment of systematic uncertainties.

The results are presented and discussed in Chapter 7 and there are final comments and conclusions in Chapter 8.



# Chapter 2

## Theory

This chapter explains the theoretical motivation for the Higgs boson, how it is produced at the LHC and its decay into two photons. The convention  $c = \hbar = 1$  is assumed everywhere. Four vector indices are labelled by  $\mu$  and  $\nu$ , whilst  $i, j, k$  are used to label  $SU(2)$  generators and  $a, b, c$  used to label  $SU(3)$  generators.

### 2.1 The Standard Model

The SM of Particle Physics is one of the crowning achievements of 20th century science. Its accuracy under high precision tests and prediction of subsequently observed phenomena in high energy physics is testament to its success. It is a gauge Quantum Field Theory (QFT) which provides a description of the fundamental particles of matter and three of the four known forces which govern their interactions; electromagnetism, the weak force and the strong force. Gravity is not included in the SM but, given the small scale nature of particle interactions and the relative weakness of gravity in comparison to the other forces, its exclusion has a negligible impact on the predictive power of the SM, at the energy scales of all experiments so far.

#### 2.1.1 Fundamental particles and forces

The fundamental matter particles described by the SM, which compose all of the known matter in the universe, are spin-1/2 fermions which obey the Dirac equation

$$(i\gamma^\mu\partial_\mu - m)\psi = 0, \tag{2.1}$$

where  $\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu}$ ,  $\eta^{\mu\nu}$  is the Minkowski metric  $(+, -, -, -)$ ,  $\partial_\mu$  is the covariant derivative and summing is implied over repeated indices[3].

The matter particles fall into two broad categories, those which interact with the strong force, the *quarks*, and those which don't, the *leptons*. When examining these particles in a table it is apparent there is some symmetry and beauty in their structure. There are six *leptons*; the electron ( $e$ ), muon ( $\mu$ ), tau ( $\tau$ ) and their corresponding neutrinos ( $\nu_e, \nu_\mu, \nu_\tau$ ), and six *quarks*; known as up, down, strange, charm, bottom and top ( $u, d, s, c, b, t$ ). Each of them has a corresponding antiparticle equal in mass but with opposite charge. The *quarks* are distinguished by their interaction with the strong-force which by its nature confines quarks to bound states. The strong force potential between two quarks contains a term linear in the distance between them,  $r$ . This overpowers the  $1/r^2$  term, analogous to that in electromagnetism, such that the more energetically favourable solution for overcoming the potential is the creation of a new quark-antiquark pair. Consequently, quarks are never observed as free states but as composite particles (*hadrons*) of two types: *mesons*, which consist of a quark-antiquark pair, and *baryons*, which consist of quark triplets. The conserved currents of the strong force, analogous to the electromagnetic charge, are colour charge which are denoted red, green or blue. One realisation of quark confinement is that free observable states are colourless; mesons contain colour anti-colour pairs and baryons contain a quark of each colour. A summary of the matter fermions along with their masses and their electromagnetic charge is given in Table 2.1.

Family	<i>Leptons</i>			<i>Hadrons</i>		
	Particle	Mass (MeV)	Charge	Particle	Mass (MeV)	Charge
I	$e^-$	0.511	-1	$u$	2.3	+2/3
	$\nu_e$	0	0	$d$	4.8	-1/3
II	$\mu^-$	105	-1	$s$	95	-1/3
	$\nu_\mu$	0	0	$c$	1.275 GeV	+2/3
III	$\tau^-$	1777	-1	$b$	4.18 GeV	-1/3
	$\nu_\tau$	0	0	$t$	173 GeV	+2/3

**Table 2.1:** The fundamental matter particles. Each particle is spin-1/2 and also has a corresponding anti-particle. All values taken from Ref[4].

The fundamental forces in the SM act via the exchange of a spin-1 vector boson. For the electromagnetic force it is the photon,  $\gamma$ , for the weak force the  $W^\pm$  and  $Z$  bosons

Force	Particle	Mass (GeV)	Charge
Electromagnetic	$\gamma$	0	0
Weak	$W^\pm$	80.4	$\pm 1$
	$Z$	91.2	0
Strong	$G^a$	0	0

**Table 2.2:** A summary of the fundamental force-carrying particles in the SM. All values taken from Ref.[4].

and for the strong the force the gluons,  $G^a$ , of which there are 8. The force-carrying particles are summarised in Table 2.2. One noticeable difference is that the photon and gluons are massless whilst the  $W^\pm$  and  $Z$  have a large mass. This becomes something of a problem when trying to unify the electromagnetic and weak forces as we somehow need to generate mass terms for the heavy bosons without doing so for the photon whilst maintaining the symmetry in the system. It turns out this can be done using spontaneous symmetry breaking which gives rise to one other particle in the SM which has not been mentioned yet, the Higgs boson. The next part of this chapter considers the symmetries involved in the SM, how they can be broken, why it is necessary that they are broken and why this necessitates the existence of a massive scalar particle.

### 2.1.2 Gauge Theories

Symmetry, and the mathematical dynamics of symmetry, are an incredibly important tool for describing fundamental physical principles. In 1918 Emmy Nöether proved that for each symmetry of the action of a physical system which can be written in the Lagrangian formalism there is a corresponding conserved quantity[5, 6]. Energy and momentum conservation are two typical examples of this which are particularly appropriate, and desirable, for particle physics. For any theory which is invariant under spatial translations (we should certainly demand that a physical principle follows the same laws anywhere in space) then Nöether's conserved quantity is momentum. For any theory which is invariant under time translations (we should also demand that a physical principle follow the same laws now, in the past and in a hundred years time) then the conserved quantity is energy. Symmetry plays a particularly important role in the SM because it is apparent that there are considerably more profound symmetries than those associated with space-time and furthermore that some of them can be broken. By demanding that any theory describing

the particle structure of the universe has the appropriate conservation properties, the dynamics of the theory can be constructed in the Lagrangian formalism by requiring that it be invariant under the relevant symmetries.

The SM is a quantised gauge theory, which is to say that the SM Lagrangian is invariant under certain local transformations. These are known as *gauge symmetries* which form a symmetry group, also called a gauge group. For each independent degree of freedom in the symmetry group there exists a generator of the group which manifests itself in the theory as a vector field, also known as a gauge field, and for a quantum theory these are spin-1 bosons. These gauge fields must be included in the mathematical formalism of the Lagrangian to ensure its invariance under the local gauge transformations[7]. This is manifested mathematically by substituting the derivative,  $\partial^\mu$ , in the Dirac equation (Eq. 2.1) for a covariant derivative

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - igA_\mu, \quad (2.2)$$

where  $A_\mu$  represents the gauge field required to maintain local invariance. In the SM we recognise these gauge fields as the force-carrying particles described in Table 2.2. This means that for each fundamental force present in the SM there must be a corresponding symmetry which has the same number of independent group generators as there are gauge bosons. The SM symmetry group is

$$SU(3) \otimes SU(2) \otimes U(1). \quad (2.3)$$

The vector field required to maintain invariance under the  $U(1)$  subgroup is labelled  $B_\mu$ , and the vector fields required to maintain invariance under the  $SU(2)$  subgroup are labelled  $W_\mu^i$ , for  $i = 1, 2, 3$ . Naively one might associate these to the SM gauge bosons in Table 2.2, however in reality nature is not as compartmentalised as this. For the proper physical description one needs to unify these two forces into the electroweak force, whose symmetry group is simply  $SU(2)_L \otimes U(1)_Y$ [8, 9, 10]. The physical states are written as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (2.4)$$

$$Z_\mu = \cos(\theta_W)W_\mu^3 - \sin(\theta_W)B_\mu \quad (2.5)$$

$$A_\mu = \sin(\theta_W)W_\mu^3 + \cos(\theta_W)B_\mu \quad (2.6)$$

$$(2.7)$$

where  $A^\mu$  is the photon field and  $\theta_W$  is known as the Weinberg angle which relates the coupling strengths of the weak,  $g_2$ , and electromagnetic,  $g_1$ , interactions:

$$\frac{g_1}{g_2} = \frac{\sin(\theta_W)}{\cos(\theta_W)}. \quad (2.8)$$

The generators for the  $SU(2)$  part of the group are  $T_i = \tau_i/2$ , where  $\tau_i$  for  $i \in 1, 2, 3$  are the Pauli spin matrices. There is one additional generator for the  $U(1)$  part of the group,  $Y$ . The corresponding conserved quantities for these symmetries are weak isospin,  $t_{1,2,3}$  and hypercharge,  $y$  which are related to the electromagnetic charge,  $Q$ , by the relation

$$Q = t_3 + y/2, \quad (2.9)$$

and the factor of 2 is chosen by convention. The remaining  $SU(3)$  sector requires 8 independent vector fields, the gluons  $G_\mu^a$  for  $a = 1, 2, 3 \dots 8$ , whose generators are given by the Gell-Mann matrices,  $\lambda_a$  for  $a \in 1, 2, \dots, 8$ , with the corresponding conserved quantity being colour charge. Thus the full covariant derivative is written as

$$D_\mu = \partial_\mu - ig_1 \frac{Y}{2} B_\mu - ig_2 \frac{\tau_i}{2} W_\mu^i - ig_3 \frac{\lambda_a}{2} G_\mu^a. \quad (2.10)$$

### 2.1.3 Quark and Lepton states

The matter particles are spin-1/2 fermions and consequently labelled by spinor fields,  $\psi$ . These can be split into their left and right-handed chiral constituents using the projection operators,  $P_L$  and  $P_R$ , such that  $\psi_L = P_L \psi$  and  $\psi_R = P_R \psi$ . The electroweak is a distinctly chiral force. Left and right-handed states transform differently under  $SU(2)$  electroweak transformations. The former are electroweak doublets and the latter are electroweak singlets and for the first generation of leptons are written as

$$\psi_1 = e_R \quad : SU(2) \text{ singlet} \quad (2.11)$$

$$\psi_2 = L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad : SU(2) \text{ doublet} \quad (2.12)$$

Note that there is no right-handed neutrino. The quarks are labelled in a similar way:

$$\psi_3 = u_{R\alpha} \quad (2.13)$$

$$\psi_4 = d_{R\alpha} \quad (2.14)$$

$$\psi_5 = Q_{L\alpha} = \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix}_L \quad (2.15)$$

where the additional index  $\alpha$  describes the quark transformations in  $SU(3)$  colour space. The convention is that whenever terms in the covariant derivative, Eq. 2.10, act on fermion terms of a different matrix form they give zero. This allows the SM fermion interaction Lagrangian to be written as

$$\mathcal{L} = \bar{\psi} i \gamma^\mu D_\mu \psi, \quad (2.16)$$

where there is an implicit sum over the fermion types,  $\psi_i$  for  $i \in 1..5 = e_R, L, u_R, d_R, Q_L$ , and a sum over the fermion generations. The kinetic term is simply written as

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = -\frac{1}{4} \left( B_{\mu\nu} B^{\mu\nu} + \text{Tr}(W_{\mu\nu} W^{\mu\nu}) + \text{Tr}(G_{\mu\nu} G^{\mu\nu}) \right), \quad (2.17)$$

where  $B_{\mu\nu}$ ,  $W_{\mu\nu}^i$  and  $G_{\mu\nu}^a$  are the field strength tensors for the three SM gauge groups, which can be expressed in terms of their corresponding vector fields by the relation

$$X_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c, \quad (2.18)$$

where  $f^{abc}$  is the structure constant of the particular group in question (one of  $U(1)$ ,  $SU(2)$  or  $SU(3)$ ) and describes the commutation relationship between the group generators.

### 2.1.4 Electroweak symmetry breaking

As presented so far the SM Lagrangian has no mass terms included in it but this is clearly in contention with our observation that most of the fundamental fermions and bosons have masses. Including mass terms by hand, of the form  $m\bar{\psi}\psi$  for the fermions and  $\frac{1}{2}m^2 B^\mu B_\mu$  for the bosons, would explicitly break the  $SU(2)$  invariance and consequently is not a good solution. The reason for this is that the electroweak is specifically a left-handed force; there is no electroweak coupling to right-handed fermions. An  $SU(2)$  doublet,  $\phi$ , is required so that one can include a mass term which looks like  $m\bar{L}\phi e_R$

and preserves invariance under  $SU(2)$ . There is a way to include a field exactly like this by spontaneously breaking the symmetry such that the Lagrangian itself is still invariant whilst its vacuum state, and hence particle spectrum, is not. It turns out that this mechanism generates a mass for the massive gauge bosons whilst leaving the photon massless and predicts the existence of a massive scalar particle. One can then introduce a coupling term for this new particle with the matter fermions which dictates the size of their masses. This is known as the Higgs mechanism.

An additional term is required in the SM Lagrangian which introduces the Higgs field, an  $SU(2)$  doublet,  $\phi$ ,

$$\mathcal{L} = T - V = (D^\mu \phi)^\dagger (D_\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad (2.19)$$

where  $T$  and  $V$  are the kinetic and potential terms respectively. We can study the particle spectrum by first finding the minimum of the potential, i.e. the vacuum state, and then expanding around this. By requiring that  $\mu^2 < 0$  and  $\lambda > 0$  it is apparent that the potential is a Mexican hat shape whose minimum is non-zero and maps out a circle in the  $SU(2)$  phase space. The vacuum state can be chosen as any one of these equivalent solutions, which lie along the circle, but the convention is to pick a direction, which anyway doesn't matter as the potential only contains terms in  $\phi^\dagger \phi$ , and define the vacuum expectation value (VEV) as,

$$\langle 0 | \phi | 0 \rangle = \begin{pmatrix} 0 \\ \sqrt{-\frac{\mu^2}{2\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.20)$$

The VEV now breaks the  $SU(2)$  invariance although the extra Lagrangian term introduced in Eq. 2.19 does not. The convention then considers small perturbations around the VEV,

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (2.21)$$

Inserting this definition of the field  $\phi$  into the Lagrangian given in Eq. 2.19, where the covariant derivative,  $D^\mu$ , is defined in Eq. 2.10 (recall that the notation used was such

that the  $SU(3)$  operators  $\lambda^a G_\mu^a$  operating on an  $SU(2)$  state, such as  $\phi$ , gave zero) gives

$$\mathcal{L} = \frac{1}{2}(\partial^\mu H)(\partial_\mu H) - \mu^2 H^2 + \frac{1}{8}g_2^2 v^2 (|W_\mu^+|^2 + |W_\mu^-|^2) + \frac{1}{8}g_2^2 v^2 \left[1 + \left(\frac{g_1}{g_2}\right)^2\right] |Z_\mu^0|^2 + \text{interaction terms} \quad (2.22)$$

Only the relevant kinematic and mass terms here have been kept and all others are simply referred to as “interaction terms”. The spontaneous symmetry breaking mechanism has given rise to:

- a scalar field,  $H$ , with mass,  $m_H = \sqrt{-2\mu^2}$ ;
- two charged gauge boson fields,  $W^\pm$ , with the same mass,  $m_W = g_2 v/2$ ;
- a neutral gauge boson field,  $Z$ , with mass,  $m_Z = m_W \sqrt{1 + (g_1/g_2)^2}$ ;
- a neutral gauge boson field,  $A_\mu$ , which has no mass term.

Consequently a mass term has been generated for the  $W^\pm$  and  $Z$  bosons whilst the photon has been left massless. There is also the prediction of the SM Higgs boson. By adding  $SU(2)$  invariant Yukawa coupling terms in the Lagrangian, the fermion masses can also be generated by the Higgs boson. These terms are of the form,

$$\mathcal{L} = k_e (\bar{L}\phi e_R + \phi^\dagger \bar{e}_R L) + \left[ k_d \bar{Q}_L \phi d_R + k_u \bar{Q}_L (-i\tau_2 \phi^*) u_R + \text{h.c.} \right], \quad (2.23)$$

where h.c. represents the hermitian conjugate of the preceding terms in brackets and  $k_{e,u,d}$  are the Higgs-fermion couplings which are directly related to the mass of the fermions by  $m_f = k_f v/\sqrt{2}$ . Neither the value of these couplings nor even the presence of such terms is determined by the gauge principle but they allow the theory to accommodate non-zero fermion masses via the Higgs boson. The relationship between the experimentally observed fermion masses and these couplings allows indirect constraints and properties of the Higgs boson to be calculated. The full SM Lagrangian can be written by summing all the terms previously discussed and simplified to

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}i\gamma^\mu D_\mu\psi + |D\phi|^2 + \mu^2|\phi|^2 - \lambda|\phi|^4 + [\psi_i k_{ij} \psi_j \phi + \text{h.c.}]. \quad (2.24)$$

This last section summarises the work of many: Higgs, Englert, Brout, Guralnik, Hagen, Kibble, Anderson, Nambu and Goldstone on spontaneous symmetry breaking



and mass emergence[11, 12, 13, 14, 15, 16, 17], Glashow, Weinberg and Salam on the electroweak mode[8, 9, 10] and t’Hooft and Veltman on the renormalisability and unitarity of the SM[18, 19]. It is quite amazing that using this model of electroweak symmetry breaking allowed Glashow, Weinberg and Salam to predict the existence, and the masses, of the  $W^\pm$  and  $Z$  bosons. These were experimentally observed by the UA1 and UA2 experiments in 1983[20, 21]. The theory also predicts the existence of a massive scalar boson known as the Higgs. Whilst its mass could not be directly predicted by the theory, various precision measurements made in the run up to LHC operation suggested its mass would be light enough to be found, if the SM was to be believed, at the energy scale of the LHC. The rest of this chapter concentrates on Higgs production at the LHC and its decay to two photons.

## 2.2 Higgs production at the LHC

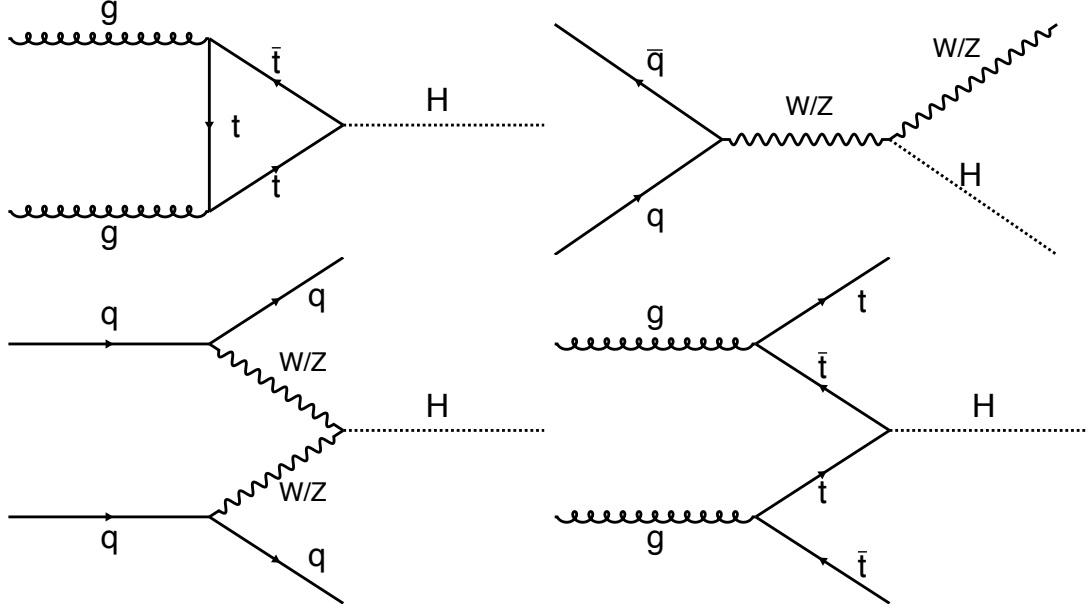
The LHC is predominantly a proton-proton collision machine capable of centre-of-mass energies far greater than any previous colliders. The dataset used for this thesis is taken at centre-of-mass energies  $\sqrt{s} = 7$  and 8 TeV. One of the experimental aims of the LHC was to find the Higgs boson or provide clues about why it cannot be seen or does not exist. Theoretical constraints, before LHC data taking, set an upper bound on the SM Higgs mass of  $m_H \leq 710 \pm 60$  GeV[22]. Previous searches by the LEP experiments obtained a lower bound of  $m_H \geq 114.4$  GeV at the 95% confidence level[23]. The LHC should be very capable of producing SM Higgs bosons anywhere in this mass range. The global electroweak precision fit for the Higgs mass prior to LHC data suggested a SM Higgs boson in the  $1\sigma$  interval  $[69 - 183]$  GeV[24].

At the LHC SM Higgs bosons are produced predominantly in one of four ways: gluon fusion ( $ggH$ ),  $VBF^1$ ,  $VH^2$  and  $t\bar{t}H$ . The Feynman diagrams for these processes are shown at leading order in Fig. 2.1. As the Higgs only couples to mass, the gluon fusion production is apparent through a top loop. This is the dominant production process, nearly 90% of Higgs bosons produced at the LHC come from gluon fusion, and provides a good measure of the Higgs coupling to fermions. The other three production modes are much smaller in cross section. However in these cases the production is in association with other particles which can be “tagged” to provide additional sensitivity

---

<sup>1</sup>sometimes also referred to as  $qqH$

<sup>2</sup>sometimes also split into  $W^\pm$  boson associated production ( $WH$ ) and  $Z$  boson associated production ( $ZH$ ) separately

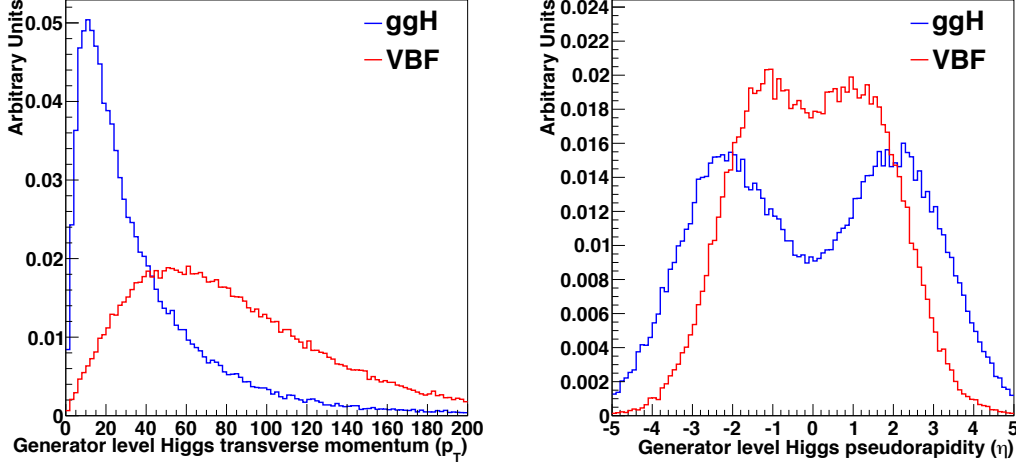


**Figure 2.1:** The four main SM Higgs production mechanisms at the LHC: gluon fusion (top left), vector boson fusion (bottom left),  $W^\pm$  and  $Z$  boson associated production (top right) and top anti-top annihilation (bottom right). The cross sections for each of these processes in proton-proton collisions is shown in Fig. 2.3.

to an analysis by reducing the background rate. The  $VBF$  production mode has a very specific topology. The associated quarks emitted in the production typically produce two high momentum but very forward jets which have a large spatial separation (in other words they are back-to-back). The Higgs produced by this mechanism also typically has a large transverse momentum. The  $VH$  production modes are associated with a  $W^\pm$  or  $Z$  boson so can be probed by searching for Higgs decays which also contain leptons and neutrinos, the latter measured at the LHC as missing transverse energy ( $\cancel{E}_T$ ). The  $t\bar{t}H$  production mode is associated to a pair of top quarks so can be probed by searching for Higgs decays containing  $b$  quarks, leptons and  $\cancel{E}_T$ , as the decay chain used for a top quark is  $t \rightarrow bW(\rightarrow l\nu)$ .

A demonstration of the differences between  $ggH$  and  $VBF$  production is shown in Fig. 2.2. The generator-level distributions of these two signals are shown as a function of the generated Higgs transverse momentum and generated Higgs pseudorapidity,  $\eta$ , defined as  $\eta = -\ln \tan(\theta/2)$ , where  $\theta$  is the polar angle measured from the beam axis.

The SM Higgs production cross section as a function of the Higgs mass,  $m_H$ , is shown for the low mass region  $90 \leq m_H \leq 300$  GeV in Fig. 2.3 for centre-of-mass energies,



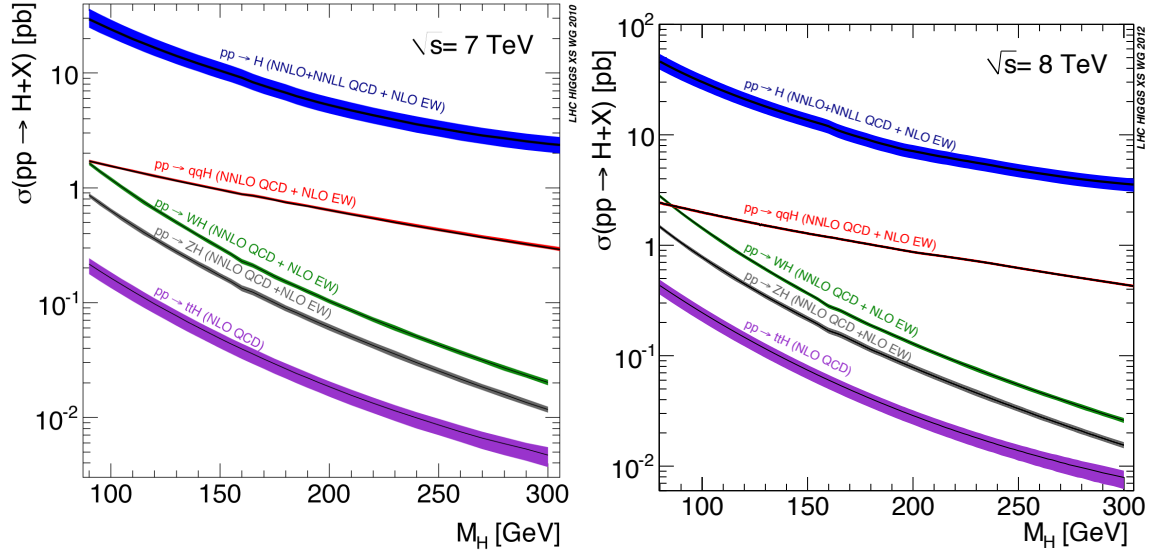
**Figure 2.2:** Generator level Higgs distribution in transverse momentum (left) and pseudorapidity (right) for production via gluon fusion (blue) and vector boson fusion (red) normalised to the same area.

$\sqrt{s} = 7$  and 8 TeV as provided by the LHC Higgs Cross Section Working Group[25]. It is clear that the production is dominated by  $ggH$  but also that this mechanism has a large theoretical uncertainty. Once the statistics of the LHC data become very high, this theoretical uncertainty becomes one of the dominant uncertainties in a cross section measurement. For a SM Higgs boson with mass  $m_H = 125$  GeV the cross section is about 18 (22) pb for  $pp$  collisions at  $\sqrt{s} = 7$  (8) TeV.

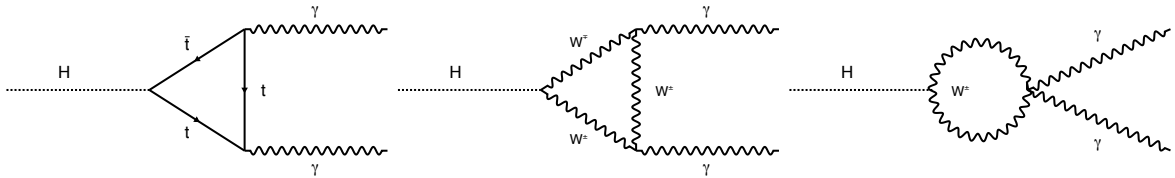
## 2.3 Higgs decay into two photons

The Higgs couplings are proportional to the mass of the coupling object. Given the photon is massless there is no direct coupling between it and the Higgs. Consequently Higgs decays to photons occur via loop diagrams with  $W$  bosons or quarks. For the latter, only the top quark loop need be considered given that the coupling is proportional to the mass and the top quark is considerably heavier than any of the other quarks. The Feynman diagrams for these processes at leading order are shown in Fig. 2.4.

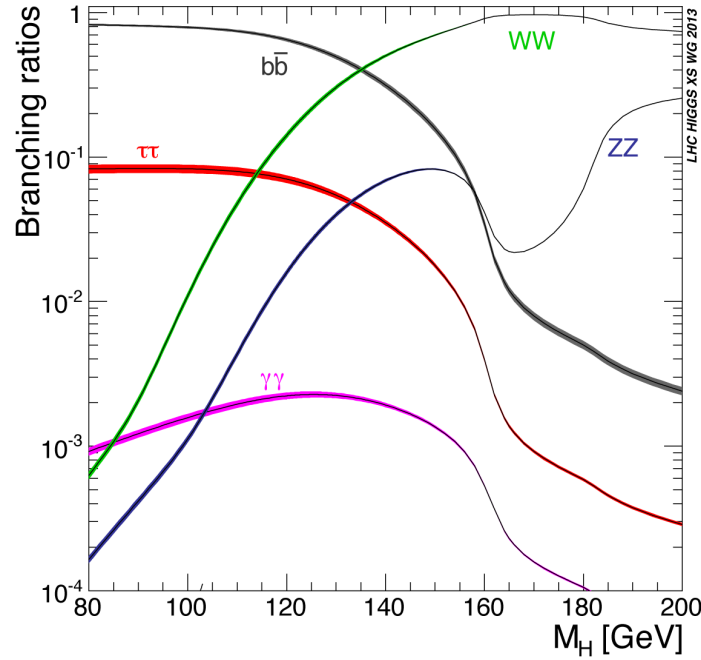
The SM Higgs branching ratio for each set of decay products is shown as a function of the Higgs mass,  $m_H$ , for the low mass region  $80 \leq m_H \leq 200$  GeV in Fig. 2.5 as provided by the LHC Higgs Cross Section Working Group[25]. The work in this thesis focuses on the Higgs decay into two photons ( $H \rightarrow \gamma\gamma$ ) whose branching fraction is shown by the



**Figure 2.3:** The SM Higgs production cross section in proton-proton collisions at the LHC for centre-of-mass energies of  $\sqrt{s} = 7$  TeV (left) and  $\sqrt{s} = 8$  TeV (right). The theoretical uncertainties on the values are shown as the coloured bands. Lines are shown for  $t\bar{t}H$  production (purple),  $ZH$  production (grey),  $WH$  production (green),  $VBF$  production (red) and  $ggH$  production (blue)[25].



**Figure 2.4:** Feynman diagrams for the SM Higgs to two photon decay at leading order.

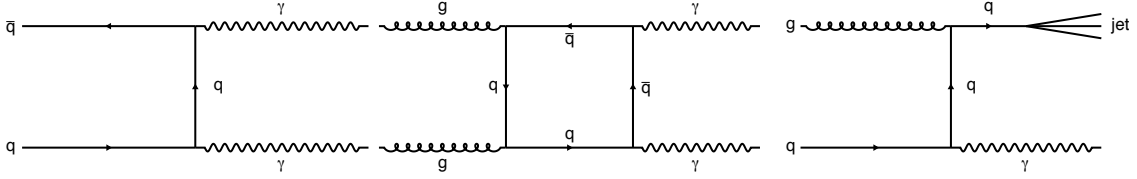


**Figure 2.5:** The branching ratio of the SM Higgs boson into various decay products as a function of the Higgs mass  $m_H$ . The theoretical uncertainties are shown as the coloured bands. Only the primary search channels are shown on this figure (Higgs decaying into  $b\bar{b}$ ,  $ZZ$ ,  $WW$ ,  $\tau\tau$  and  $\gamma\gamma$ ) for ease of viewing. There are several other possibilities left out (Higgs decaying into  $gg$ ,  $c\bar{c}$ ,  $Z\gamma$ ,  $\mu\mu$ ). This thesis concentrates on the  $H \rightarrow \gamma\gamma$  decay shown as the pink line[25].

pink line. It is apparent that  $H \rightarrow \gamma\gamma$  decays are rare. There is only a small window of Higgs masses in which  $H \rightarrow \gamma\gamma$  decay is even feasible ( $m_H < \sim 185$  GeV) and the peak of the branching fraction ( $120 < m_H < 130$  GeV) only allows a SM  $H \rightarrow \gamma\gamma$  decay 0.2% of the time. Given that the LHC Run 1 dataset used in this thesis consists of  $5.1 \text{ fb}^{-1}$  at  $\sqrt{s}=7$  TeV and  $19.7 \text{ fb}^{-1}$  at  $\sqrt{s}=8$  TeV one can expect about half a million SM Higgs bosons to be produced (assuming a value of  $m_H = 125$  GeV) of which only about 1000 decay into two photons.

### 2.3.1 Backgrounds to the $H \rightarrow \gamma\gamma$ decay at the LHC

Aside from the very low signal rate of Higgs decays to two photons, further complications arise by considering the incredibly high rate of the background processes for two photon production in proton-proton collisions. A pair of real (prompt) photons is predominantly produced by quantum chromodynamics (QCD) interactions from a proton-proton initial state via two diagrams; the so-called Born ( $q\bar{q} \rightarrow \gamma\gamma$ ) and the box



**Figure 2.6:** The prompt-prompt and prompt-fake Feynman diagrams contributing to the  $H \rightarrow \gamma\gamma$  background via the Born mode (left), box mode (middle) and  $\gamma$ +jet mode (right).

( $gg \rightarrow \gamma\gamma$ ), collectively known as *prompt-prompt* background. These two backgrounds are referred to as irreducible as they fake signal with two real photons. By using the specific kinematics of Higgs decays these backgrounds can be somewhat suppressed but the main challenge of the analysis is estimating the contamination of these processes in the signal region. The other type of background arises from final state neutral hadrons faking photons. Predominantly these are  $\pi^0$ s decaying into two almost collinear photons which fake the single photon signal. This can happen in association with one real photon,  $\gamma$ +jet (known as *prompt-fake*), or where both photons are faked by jet signals (known as *fake-fake*). Nearly all of the *fake-fake* background can be removed using the analysis techniques described in Chapters 4 and 5 such that the final analysis consists of about 70% prompt-prompt, 30% prompt-fake and  $< 1\%$  fake-fake. The Feynman diagrams for Born, box and  $\gamma$ +jet production are shown in Fig. 2.6.

One of the most important variables used in the analysis is the invariant mass of the two photons, which for signal is equivalent to the reconstructed Higgs mass. The invariant mass distribution for the background is expected to be a smoothly falling continuum whereas the signal is expected to be a narrow resonance centred at the Higgs mass. The difference between these is heavily exploited in the analysis. The diphoton invariant mass is reconstructed using,

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \alpha)}, \quad (2.25)$$

where  $E_1$  and  $E_2$  are the energies of the two photons, and  $\alpha$  is the angle between them.

The remainder of this thesis will concentrate on the details of an analysis of Higgs events decaying into two photons at the CMS experiment at the LHC. Many of the features discussed in the last section will be exploited using sophisticated computing, analysis and statistical techniques which ultimately culminate in a standalone observation of a resonance near 125 GeV and subsequent measurements of this particle's properties.

# Chapter 3

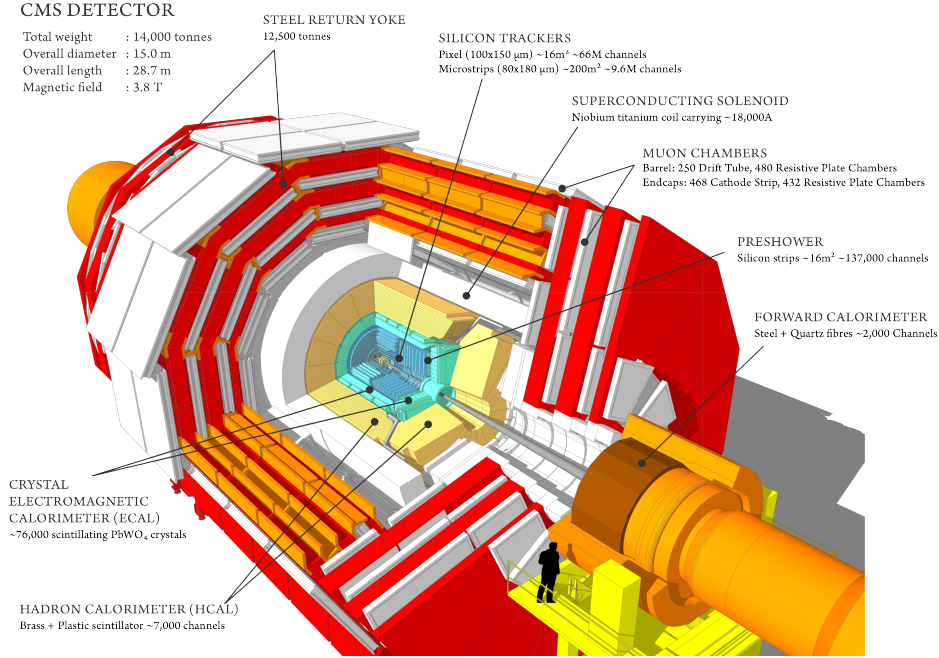
## The CMS experiment

### 3.1 The LHC

The LHC is an octagonal 27 km ring (large) proton-proton (hadron) particle collider. Using a multistage acceleration process two beams of protons are circulated in opposite directions at a centre-of-mass energy in excess of  $\sqrt{s}=7$  TeV. The protons are sourced from hydrogen and circulate in the collider in bunches.

The LHC ring consists of eight straight line segments, in which the charged beams are accelerated by an oscillating electric field using radio frequency (RF) cavities, and eight arc segments, which are filled with superconducting magnets whose effect is to circulate the beam of protons around the ring. The spacing between bunches is 50 ns for the data used in this thesis, although the LHC can operate at a bunch spacing of 25 ns, and there are about 100 billion protons per bunch. Precision magnetic fields can control the position and intensity of the beams such that they can be focused into a small space (around 64 microns at the interaction point) and collided together. The result is about 20 collisions per crossing. For the data used in this thesis the LHC operated at centre-of-mass energies of  $\sqrt{s} = 7$  and 8 TeV.

There are four points around the ring where the beams can be forced to intersect producing high energy proton-proton collisions. Particle detectors are constructed around these points such that the collision can be reconstructed with the purpose of measuring physical properties and processes, calibrating the detectors with already known processes and searching for new physics. The remainder of this chapter concentrates on a description of one of these detectors, CMS, which the author has worked on.



**Figure 3.1:** Schematic representation of CMS showing the layered structure of subdetectors; tracking system, calorimeters, magnet and muon system[27].

## 3.2 The CMS detector

The CMS detector, pictured in Fig. 3.1, is a multipurpose experiment designed for the measurement of and search for a multitude of different processes. We will primarily discuss its function as a Higgs finding machine. A more detailed description can be found in Ref.[26]. It has a cylindrical shape consisting of a barrel segment, 21.6 m long, and two endcaps, 14.6 m in diameter, aligned along the beam direction with its centre at the beam interaction point. The endcaps are nearer the beam line and so the materials in these components typically have to be able to withstand higher amounts of radiation and therefore tend to have worse performance. Many of the features of CMS exploit what one would expect for measuring Higgs decays: it has almost full coverage of the area around the collision point so that nearly every particle emanating from the collision can be reconstructed and it has many complementary subsystems (or layers) designed to measure different specific particles so that Higgs bosons can be detected through a multitude of decay modes.



For a Higgs with an intermediate mass (100 – 200 GeV) the high resolution (narrow peak) channels are  $H \rightarrow ZZ^{*1} \rightarrow l^+l^-l^+l^-$  and  $H \rightarrow \gamma\gamma$  so good energy resolution and identification of electrons and muons is desirable down to very low  $p_T$  ( $\sim O(10\text{GeV})$ ) as well as good resolution and identification of high energy photons.

The central design feature of CMS is the very powerful superconducting magnet which produces an axial magnetic field of 4T. The size of this field, as well as the density of the calorimeter materials, allows for a compact and economical design (much more so than its sister detector, ATLAS). Outside of the magnet lie the muon stations which also serve as a return yoke for the magnetic field. The muon chambers in the barrel consist of alternating layers of drift tubes and resistive plate chambers which provide both accurate timing and hit location, in order to reconstruct muons down to low energies. In the endcap the drift tubes are replaced with cathode strip chambers. Combining information from the muon subsystem with information from the inner tracking system (described below) allows muons at CMS to be reconstructed down to  $p_T = 10$  GeV with a resolution of  $\sim 1\%$ . The other three main subsystems at CMS, the tracking system and the two calorimeters, are located inside the magnet

The first layer, moving outwards from the interaction point, is the tracking system which is used to reconstruct the momentum of any outgoing charged particles and to locate the primary and secondary vertices. This is surrounded by the calorimeters, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The first is a single layer of dense, transparent crystals which collects deposits of energy left by electrons and photons which shower inside the material. The second complements this by providing a measurement of the energy deposited by hadrons (reconstructed as objects known as jets) through nuclear interactions. The HCAL is a sampling calorimeter in which the active material (plastic scintillator) is sandwiched between a dense absorbent material (brass or steel). This extends the radiation length of the calorimeter (clearly accommodating the compact design) and provides pointing information but degrades the resolution of reconstructing jets.

CMS uses a right-handed Cartesian coordinate system with the origin at the interaction point and the  $z$ -axis pointing along the beam line. The  $x$ -axis points towards the centre of the LHC ring and the  $y$ -axis points vertically upwards. The azimuthal angle,  $\phi \in [-\pi, \pi]$ , is defined with respect to the  $x$ -axis in the transverse ( $x - y$ ) plane. The polar angle  $\theta$  is measured from the  $z$ -axis. Commonly, the direction of an outgoing particle is defined by

---

<sup>1</sup>A \* denotes that one  $Z$  can be off mass shell

$\phi$  and its pseudo-rapidity  $\eta$ ,

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \quad (3.1)$$

The LHC is capable of producing 40M bunch collisions per second, and each can result in several  $p$ - $p$  collisions. However many of these are not hard interactions, the result being that the outgoing particle debris follows the beam line. A hard (and therefore interesting) collision is characterised by the amount of energy produced in the transverse ( $x - y$ ) plane. Therefore particles are commonly characterised by the projection of their momentum onto this plane, their transverse momentum,

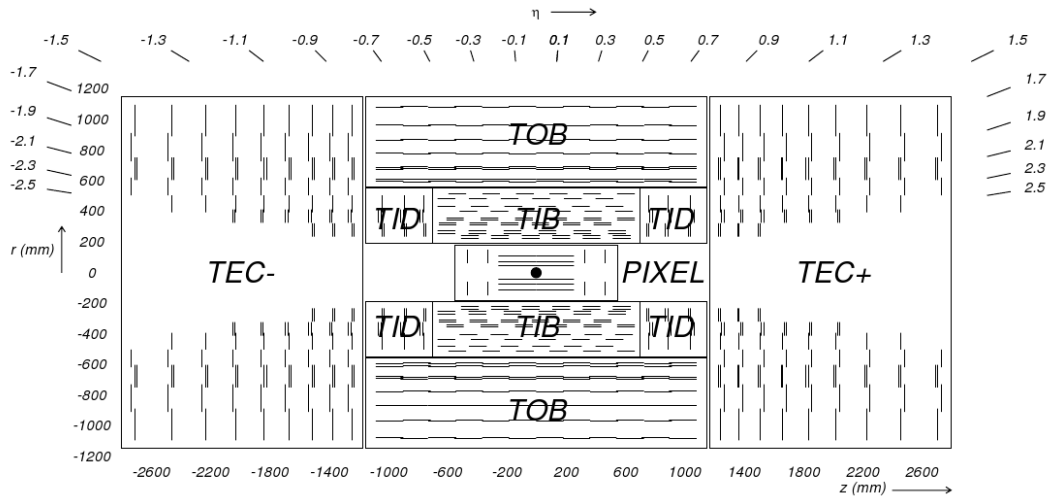
$$p_T = \sqrt{p_x^2 + p_y^2}, \quad (3.2)$$

and the corresponding transverse energy,  $E_T = E \sin(\theta)$ .

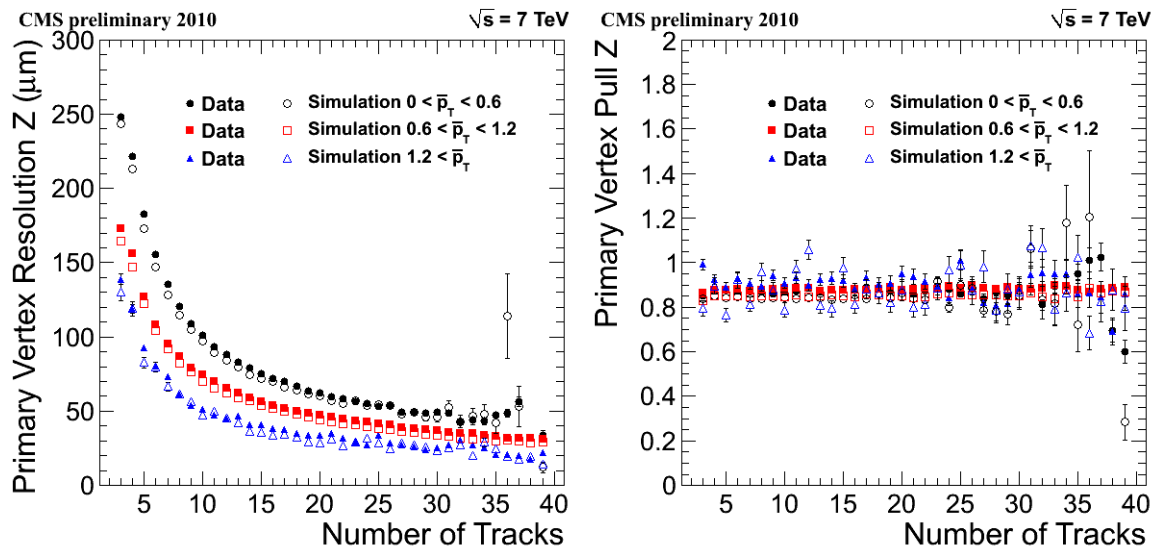
### 3.2.1 Tracking system

During nominal LHC running conditions in 2012 there are on average over 1000 particles from up to 50 overlapping  $p$ - $p$  collisions (pileup) per bunch crossing (every 50 ns). The tracker is designed to efficiently and precisely reconstruct all charged particle trajectories, and thus their position and momentum, which are known as tracks. Due to the vast number of tracks emanating from multiple vertices in typical LHC collisions, the tracking material and electronics are required to have high granularity, a fast response and be radiation hard. This conflicts with another important design feature of the tracker which is the aspiration to use the minimal amount of material in order to reduce multiple scattering, bremsstrahlung, photon conversion and nuclear interactions before particles reach the calorimeters. These criteria motivate the choice of silicon throughout the CMS tracking system. The structure of the CMS tracker is shown in Fig. 3.2 and consists of a central pixel detector (PIXEL) surrounded by layers of silicon strips aligned parallel to the beam line in the barrel (TIB and TOB) and perpendicular to the beam line in the endcap (TID and TEC). The total active surface of the tracking system is about 200 m<sup>2</sup> and is instrumented with  $\sim 10^7$  channels[28].

By making multiple precise measurements of tracks as they pass through the pixel and silicon layers the track trajectories can be reconstructed and their momentum calculated using their curvature in the  $\phi$  plane due to the axial magnetic field. Tracks are grouped

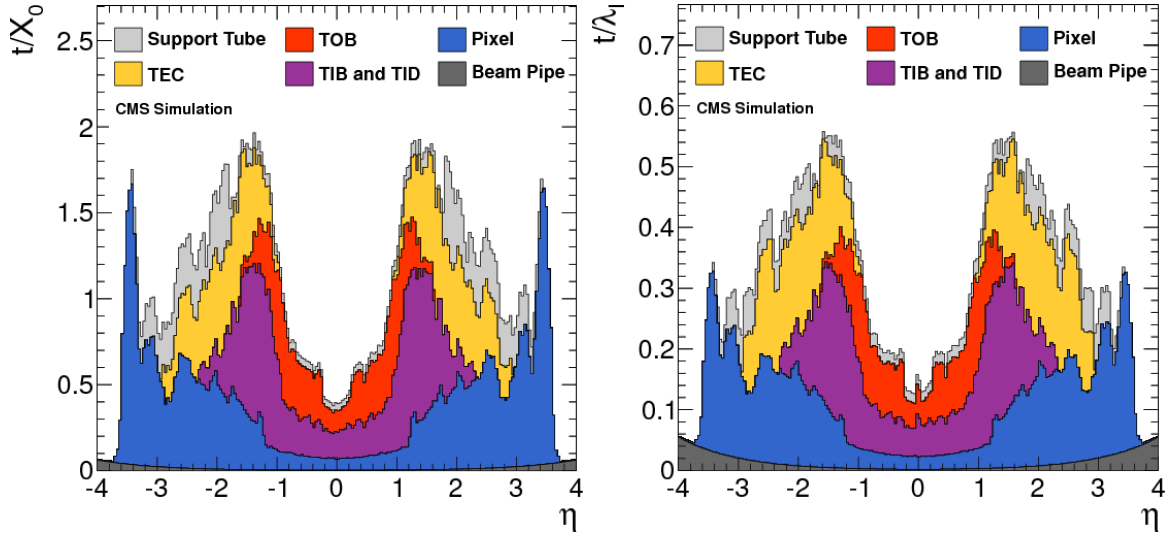


**Figure 3.2:** A diagram of the CMS tracking system showing the PIXEL detector and outer silicon strip layers[26].



**Figure 3.3:** Primary vertex position resolution and pull in  $z$  as a function of the number of tracks originating from that vertex. The different colours represent three different bins in the average track  $p_T$  where data is shown as solid points and simulation as open points[29].

together (requiring that their separation is less than 1 cm in the  $z$  coordinate at the point of closest approach to the beamline) and assigned to a common point or origin (the primary vertex). The vertex resolution is driven both by the number of tracks originating from a particular vertex and how large their average  $p_T$  is. This is shown in Fig. 3.3 for preliminary data taken in 2010 at  $\sqrt{s} = 7$  TeV.

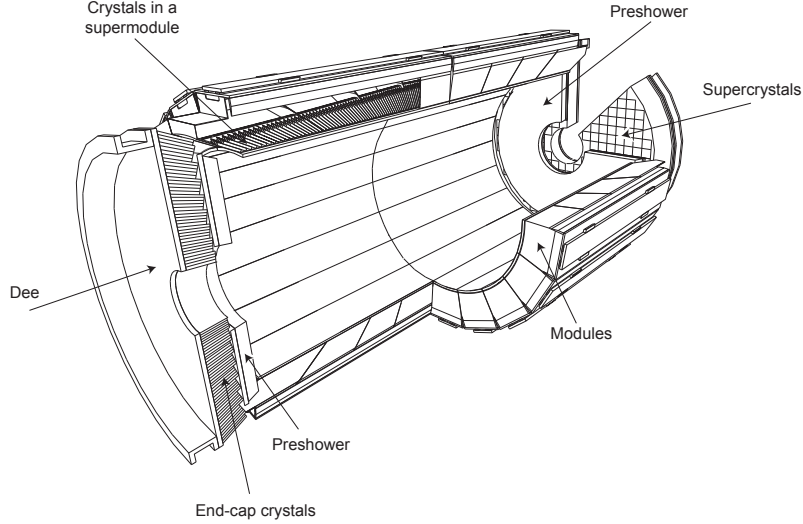


**Figure 3.4:** The amount of CMS tracker material in radiation lengths ( $X_0$ ) on the left and in nuclear interaction lengths ( $\lambda_I$ ) on the right as a function of  $\eta$  for the different tracking subsystems[30].

The amount of material in the tracking system is shown in Fig. 3.4 which demonstrates as a function of  $\eta$  which subsystems of the tracker, beam pipe and services contribute to the material in between the interaction point and the calorimeters in radiation lengths ( $X_0$ ) and nuclear interaction lengths ( $\lambda_I$ ). As shown in Fig. 3.2 the tracker has full coverage in  $\phi$  and for  $|\eta| \leq 2.5$ . As we will see later the tracking system is very important for the  $H \rightarrow \gamma\gamma$  search at CMS as without it locating the primary vertex would be practically impossible.

### 3.2.2 Electromagnetic calorimeter

The ECAL is used to reconstruct the energy of electrons and photons which deposit their energy via electromagnetic showers in the calorimeter material. The shower inside the crystal produces photons whose total energy is proportional to the energy of the incoming particle. Hence the light output from the shower, which can be measured by photodiodes at the back of each crystal, in turn provides a measurement for the energy of the original particle. The ECAL has almost full hermetic coverage of the interaction point and consists of a single layer of lead tungstate ( $\text{PbWO}_4$ ) crystals. The crystals are laid out in a quasi-projective geometry such that they point towards the interaction point with an offset of  $3^\circ$  making it much less likely that a photon, or electron, will pass straight through a gap between crystals. The ECAL consists of a barrel section and two endcap



**Figure 3.5:** Schematic drawing of the CMS ECAL layout showing the “modules” of crystals in the ECAL barrel and the “Dees” of crystals in the ECAL endcap[26].

disks which are preceded by a preshower (to aid with  $\pi^0$  rejection); a schematic drawing is shown in Fig. 3.5. The fiducial region excludes the region between the barrel and endcap (to prevent reconstruction of showers which overlap both subsystems) yielding an ECAL coverage of  $|\eta| < 2.5$ , but not in the range  $1.444 < |\eta| < 1.556$ , and full coverage in  $\phi$ . Typical crystal dimensions are  $2.2 \times 2.2$  cm at the ECAL face, with a depth of 23 cm (which amounts to  $25.8 X_0$ ) implying that practically the whole depth of the shower is contained within this single layer[31].

$\text{PbWO}_4$  has some attractive properties for an electromagnetic calorimeter especially when considering  $H \rightarrow \gamma\gamma$  decays. In order to achieve good energy resolution it is desirable to have a design in which most of the electromagnetic shower from an incoming photon or electron is contained within a single crystal, as less of the shower (and therefore energy) is lost in cracks and gaps in between crystals.  $\text{PbWO}_4$  has a short radiation length ( $X_0 = 0.89$  cm), which complements a compact design as the entire depth of the shower can be contained in a crystal which is not very long, and it has a small Molière radius (1.96 cm) which means that the lateral size of the shower is small. This generally means that the cross-sectional size of a crystal can be small (yielding high granularity of the detector) whilst still containing a large percentage of the shower. It has a very short scintillation time decay constant (85% of the light is collected in 25 ns), in other words it is very “fast”, allowing the energy in the shower to be collected and measured very quickly. This is clearly desirable in an environment like the LHC when collisions are happening up to every 25 ns. One drawback of  $\text{PbWO}_4$  (apart from expense) compared to other crystal

materials is its relatively low light yield at room temperature ( $\sim 50$ -80 photons/MeV). This is overcome at CMS with the use of silicon avalanche photodiodes (APD) in the ECAL barrel and vacuum phototriodes (VPT) in the endcap, which amplify the signal enough to make accurate measurements of the energy of the shower.

### Energy resolution

As we have seen previously the diphoton invariant mass is given by,

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos\alpha)}, \quad (3.3)$$

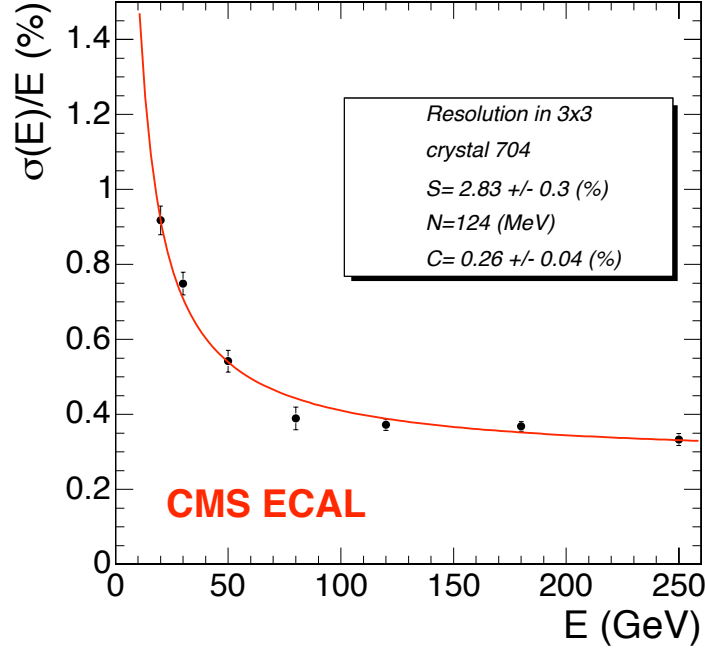
where  $E_1$  and  $E_2$  are the energy of the two photons and  $\alpha$  is the angle between them. Therefore the mass resolution has terms that depend on the photon energy resolution and angular resolution,

$$\frac{\sigma_M}{M} = \frac{1}{2} \left[ \frac{\sigma_{E_1}}{E_1} \oplus \frac{\sigma_{E_2}}{E_2} \oplus \frac{\sigma_\alpha}{\tan(\alpha/2)} \right], \quad (3.4)$$

where  $\sigma$  denotes the resolution and  $\oplus$  the quadratic sum. It is therefore desirable to have both good energy resolution and good position resolution for photons (accurate position measurements at the ECAL face alongside knowledge of the primary vertex can be used to calculate the individual photons' direction and ergo the angle between them). The energy resolution is usually then further parametrised as,

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (3.5)$$

where  $a$  is the stochastic term,  $b$  the noise term and  $c$  a constant term. In order to achieve the best possible resolution, all three of these terms need to be of a similar order and as small as possible. The size of these terms has been determined from test beam data in Ref[26]. The stochastic term is driven by the material choice and detector type so cannot be improved once the machine is built. The main contributions to this term are lateral shower containment fluctuations, photostatistics and fluctuations in the energy deposited in the preshower absorber. For a homogeneous calorimeter (the CMS ECAL is not a sampling calorimeter) made of  $\text{PbWO}_4$  the size of this term is good ( $a = 0.028 \pm 0.003 \text{ GeV}^{\frac{1}{2}}$ ). The constant term, which depends on non-uniformity of longitudinal light, intercalibration errors and energy leakage from the back of the crystal, can be minimized by use of *in situ* calibration of individual crystals and amounts

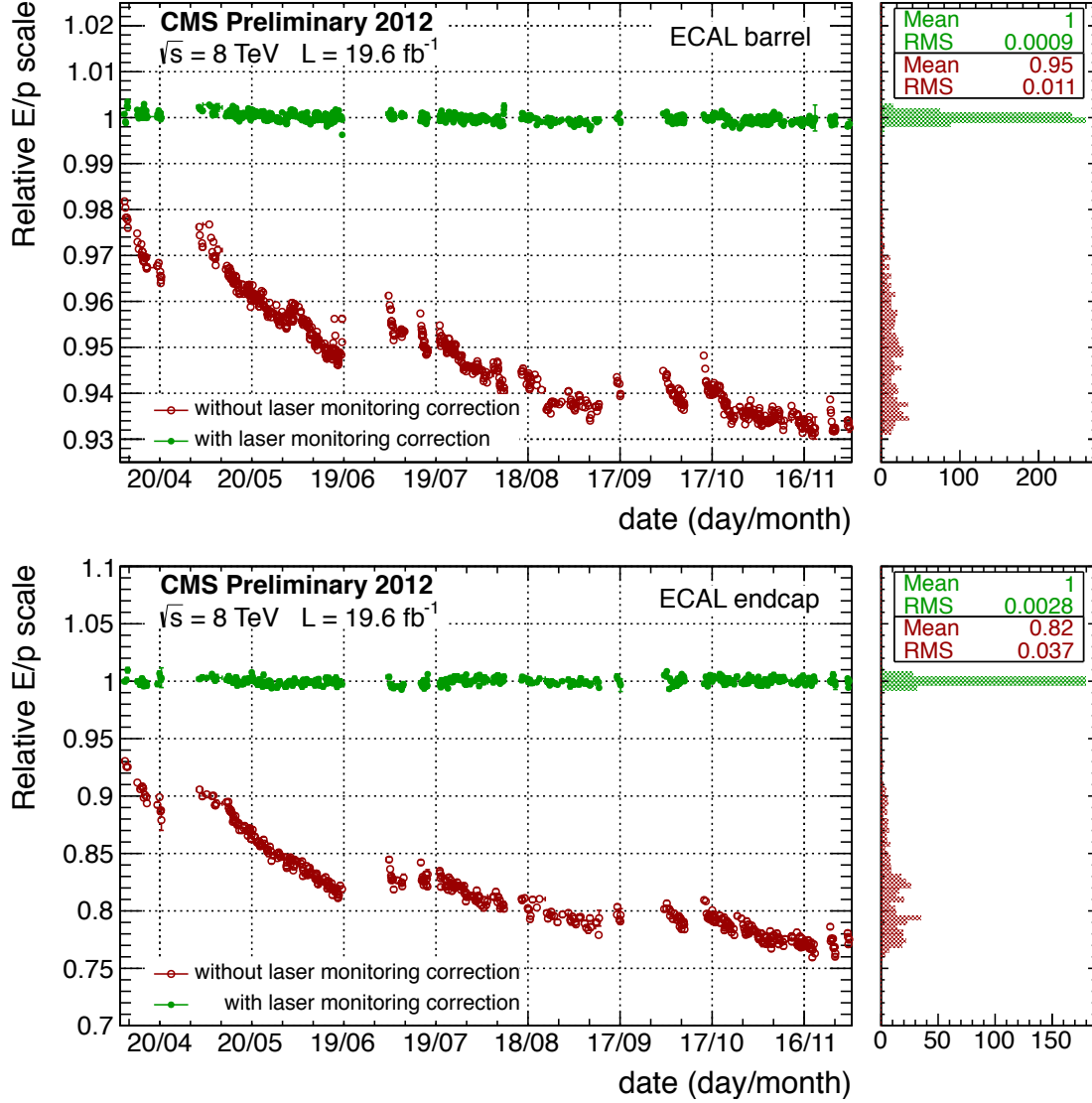


**Figure 3.6:** The ECAL energy resolution,  $\sigma_E/E$ , as a function of electron energy measured from test beam data. The energy is measured in a  $3 \times 3$  array of crystals centered on the crystal of electron impact[26].

to  $c = 0.26 \pm 0.05\%$ . The noise term which has contributions from electronics noise (including signal digitisation) and event pileup (additional particles causing overlapping signals) is measured as  $b = 126$  MeV. The ECAL energy resolution,  $\sigma_E/E$ , as a function of electron energy is shown in Fig. 3.6 as measured from a beam test[26].

### Transparency corrections

Due to the high particle flux present at CMS the ECAL crystals and electronics have to be radiation hard, especially in the endcaps. This is another motivating factor for using  $\text{PbWO}_4$  as the crystal material. The crystals are additionally doped with Nb to improve the induced absorption coefficient. Over time and long exposure to radiation the crystals lose their transparency, although there is considerable natural recovery during down periods. An important part of the ECAL monitoring and calibration comes in the form of transparency corrections to compensate for these losses. At regular intervals during LHC running laser pulses are injected into the crystals to measure the crystal response. Two different wavelengths of laser are used, one blue ( $\lambda = 440$  nm) which is very similar to the scintillation emission peak and therefore expected to be affected by transparency



**Figure 3.7:** Ratio of the electron energy,  $E$ , to the electron momentum,  $p$ , measured in the CMS barrel (top) and endcap (bottom) for 2012 data. The open red (solid green) points show the performance before (after) the laser monitoring derived corrections[32].

changes in a similar way to typical scintillation light, and one red ( $\lambda = 796$  nm) which is far from the scintillation emission peak and affected very little by the changes in transparency. Hence, by comparing the red and blue laser light response, time and  $\eta$  dependent corrections for crystal transparency loss can be calculated. A closure test for these corrections, in 2012 data, is shown in Fig. 3.7 which shows the ratio of electron energy (calculated from the ECAL) to electron momentum (calculated from the tracker) before and after transparency (or laser) corrections.



## Preshower

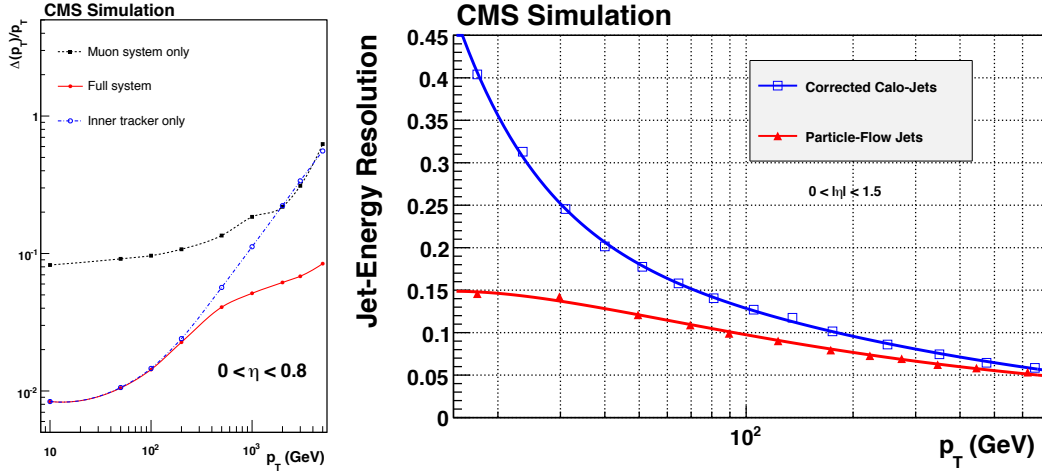
The dominant source of background to high energy photon signals are neutral mesons, mainly pions ( $\pi^0$ ), which sometimes decay into two approximately collinear photons and can therefore look very much like a single high energy photon. The ECAL endcap is preceded by a preshower to specifically target this and provide the endcap with a higher granularity. The higher granularity helps to discriminate  $\pi^0$  decays from direct photons using the spatial separation of the two photons from the  $\pi^0$ . The preshower is a sampling calorimeter which consists of two layers: a lead plate, to initiate the shower, in front of a fine grained silicon detector which has two layers of orthogonal strips. There are other characteristics of  $\pi^0$  decays which can help in differentiating them from real (“prompt”) photons, these include isolation (discussed later in this Chapter in Sec. 3.5) and the shower shape (discussed in Chapter 5 in Secs. 5.1.1 and 5.1.2).

### 3.2.3 Hadronic calorimeter

Surrounding the ECAL, but still inside the magnet, is a sampling HCAL which has geometric coverage up to  $|\eta| < 5.0$  when including the specialised forward components. It consists of alternating layers of brass plates and plastic scintillators (where in the very forward region the brass is replaced with steel). The HCAL thickness constitutes around 10-15 nuclear interaction lengths depending on  $\eta$ . Any outgoing hadrons from the interaction (of which there are many for a typical event with high  $E_T$ ) get reconstructed as objects known as “jets” by amalgamating information from the tracking system, the ECAL and the HCAL. This process, in which individual hadrons are first reconstructed and then collected to form jets, is described in more detail in the particle flow section below (section 3.4).

### 3.2.4 Muon Chambers

Outside of the magnet lie the CMS muon chambers. These consist of alternating layers of drift tube chambers (cathode strip chambers) in the barrel (endcap) and resistive plate chambers which also act as a return for the magnetic flux. The muon detector has coverage up to  $|\eta| < 2.4$  and given the particularly conspicuous signature of muons (several hits in the tracker and hits in each muon station layer) the reconstruction efficiency and momentum resolution of muons is very good even down to low  $p_T$ . The



**Figure 3.8: Left:** The muon  $p_T$  resolution, in MC simulation, as a function of  $p_T$  for muons in the range  $0 < |\eta| < 0.8$  when using only the muon system (black), only the tracking system (blue) and using both (red)[26].

**Right:** The jet energy resolution, in MC simulation, as a function of jet  $p_T$  when using PF jets as compared to calo jets[33].

muon resolution as a function of  $p_T$  is shown in the left hand plot of Fig. 3.8 for simulated data at  $\sqrt{s} = 7$  TeV.

### 3.3 Photon reconstruction

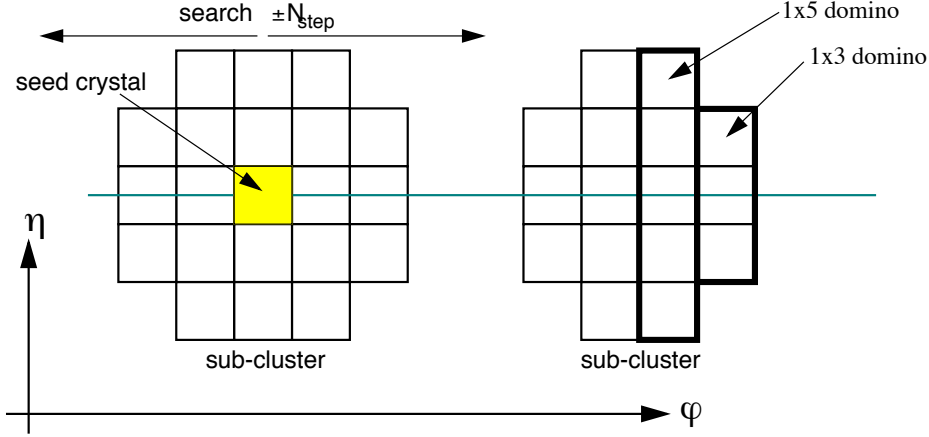
Calculating an incoming photon’s energy amounts to summing the energy deposited by the electromagnetic shower which is initiated by the photon impact at the crystal face. Due to the presence of material in the beam pipe and tracking system (see Fig. 3.4) about 40% of photons will convert into an electron-positron pair before they reach the ECAL. If this is the case the shower will spread out in  $\phi$  due to the presence of the magnetic field; the photon converts to electrons which bend and bremsstrahlung-radiate additional photons, and can be distributed among multiple (up to hundreds of) crystals. Consequently clustering (pattern matching) algorithms are deployed to calculate the “raw” photon energy. Corrections to this energy are subsequently applied to account for any energy loss as explained in Sec. 4.3. The shower will appear as a local maximum amongst a spatial neighbourhood of crystal energy deposits and so the algorithms used search first for the most energetic crystals (known as the “seed” crystals) and then extend to amass as a large a fraction as possible of the original shower energy. There are three cases to include which are i) photons which reach the ECAL without interacting with

any of the intermediate material in the beam pipe and tracking system (referred to as unconverted photons), ii) photons which convert into an electron-positron pair inside the tracker and shower in the barrel, iii) photons which convert and shower in the endcap. We will first consider the case of converted photons in the barrel. This is so similar to the case of a real electron that the identical algorithm is used for electrons as well. The method used is known as the “Hybrid” algorithm, depicted in Fig. 3.9, which makes collections of clusters known as a “supercluster”: a cluster being a set of crystals which pick up an electron or a photon produced by bremsstrahlung and a collection of clusters being a set of these which make up all the electrons and photons radiated from the original object. The algorithm can be described as a five step process as follows:

1. locate the seed crystal which is the maximum energy crystal in the search region, not already in a cluster, and which must satisfy the threshold condition,  $E_T > 1$  GeV;
2. extend in  $\eta$  to construct a “domino” which is  $1 \times 3$  crystals in  $\phi \times \eta$ . If the  $E_T$  of the central crystal in the  $1 \times 3$  domino is greater than 1 GeV then extend this to a  $1 \times 5$  domino in  $\phi \times \eta$ . Note that for the first domino this condition is always true but as the algorithm extends out in  $\phi$  (see following steps) it is not always the case that the central crystal of a domino has  $E_T > 1$  GeV;
3. traverse along  $\phi$ , up to a maximum of 17 crystals in both directions, adding dominoes in this way. If a domino has less energy than 0.1 GeV then it is excluded;
4. of the clusters of adjacent dominoes the seed domino (most energetic) must have energy greater than 0.35 GeV;
5. repeat, starting from Step 1, to build a new supercluster.

In this way a collection of dominoes are clustered in  $\phi$  creating a “supercluster” of smaller clusters (which has a maximum area of  $5 \times 35$  in  $\eta \times \phi$ ). A similar but slightly modified algorithm is used for photons and electrons in the endcap. This is known as the “Multi $5 \times 5$ ” algorithm and proceeds as follows:

1. locate the seed crystal which is the maximum energy crystal in the search region, not already in a cluster, and which must satisfy the threshold condition,  $E_T > 180$  MeV;
2. the  $5 \times 5$  array of crystals surrounding the seed are summed to make a cluster;
3. crystals at the edge of the cluster can seed new overlapping  $5 \times 5$  clusters if they are local maxima compared to their neighbouring crystals;



**Figure 3.9:** The domino construction setup of the hybrid clustering algorithm[34].

4. proceed in this way in any direction building an overlapping collection of  $5 \times 5$  clusters to create a supercluster;
5. repeat, starting from Step 1, to build a new supercluster.

The final case concerns photons which reach the ECAL without converting. These have a shower that is much more localised in  $\eta$  and  $\phi$ ; about 94% of its energy is deposited in an area of  $3 \times 3$  crystals and more than 97% in an area of  $5 \times 5$  crystals. This provides a definition for the conversion variable,  $R_9$ , such that,

$$R_9 = \frac{E_{3 \times 3}}{E_{SC}} \begin{cases} \text{unconverted if } R_9 \geq 0.94 \\ \text{converted otherwise.} \end{cases} \quad (3.6)$$

where  $E_{SC}$  is the energy of the supercluster. If a photon fulfills the unconverted requirement given above (Eq. 3.6) the photon energy is reconstructed as the sum of all energy deposited in the  $5 \times 5$  array of crystals which surround the most energetic crystal. It has been shown that using a fixed window for non-converting photons yields a better energy resolution than any clustering procedure[34].

The location of a supercluster is determined as the energy-weighted mean position of the crystals in the supercluster. This gives a position resolution which is much smaller than the size of an individual crystal ( $\sim 20 \times 20 \text{ mm}^2$ ). Values for an electron of  $p_T = 35 \text{ GeV}$  in the ECAL barrel, in the absence of pileup, are  $\sigma_\eta = 1.0 \times 10^{-3}$ ,  $\sigma_\phi = 1.6 \text{ mrad}$ .

### 3.3.1 Electron and photon differences

It is clear that when only considering the ECAL there is little difference between electrons and photons. Consequently at the level of the calorimetry there is no distinction between them, simply the idea of a supercluster which can apply to both. In the case of an electron, information from the tracker can be included using a Gaussian sum filter algorithm[35], where a series of compatible track hits are associated to the supercluster. This is used to provide a supplementary measurement of the electron's momentum and consequently improve the energy resolution of electrons. When considering photons for an analysis an electron veto must be applied requiring that no track hits should be found close to the interaction point near the photon direction (see Chapter 4). An important feature of supercluster reconstruction at CMS is that when all track information is ignored electrons and photons are identical. This is a principal ingredient in the  $H \rightarrow \gamma\gamma$  analysis which allows data driven calibration, validation and efficiency measurements of photons using electrons from  $Z$  decays (see Sec. 4.6).

## 3.4 Particle flow and jets

A traditional approach to detector-based reconstruction is to consider the objects we measure as opposed to the underlying physics objects. These are often known as calorimeter objects, for example, a track, an electromagnetic shower or a calorimeter jet. A more modern approach is to couple information in all of the subdetector systems together to reconstruct more physical objects. For example, a charged hadron will leave a track, deposit some energy in the ECAL and deposit the rest of its energy in the HCAL. This technique of reconstruction is known as PF<sup>2</sup> and is particularly useful when considering jets. Whilst an electron, photon or muon have fairly characteristic signatures, hadrons often do not. The abundant number of gluons and quarks, produced in LHC collisions, hadronise via the strong interaction as they travel away from the interaction. As they have typically high momentum the hadronisation occurs in a collimated fashion leaving a signature of several tracks and an HCAL cluster. The PF reconstruction algorithm can be simplistically viewed as the following procedure (more details given in Ref[36]):

---

<sup>2</sup>The official name inside CMS is global event description (GED) although this is rarely used

1. make small clusters from each subdetector component; tracks, ECAL and HCAL clusters to create a list of unassociated objects;
2. match tracks and clusters together and associate them to a newly reconstructed particle known as a PF candidate:
  - tracks and clusters associated with hits in the muon chambers are tagged as *muons* and removed from the list;
  - tracks and clusters associated with electrons, including bremsstrahlung photons, are tagged as *electrons* and removed from the list;
  - tracks associated to an HCAL cluster are tagged as *charged hadrons*, assigned an energy ascertained from a weighted average of the cluster energy and track momentum and subsequently removed from the list;
  - any excess cluster energy in the HCAL is assigned as a *neutral hadron* and removed from the list;
  - if an ECAL cluster is associated to an HCAL cluster and a track, it is assigned as a *charged hadron* with the appropriate weighted energy and removed from the list;
  - if an ECAL cluster is associated to an HCAL cluster with no track, it is assigned as either a *photon* or a *neutral hadron* depending on the HCAL to ECAL energy ratio and removed from the list;
  - any remaining unlinked candidates are assigned as *photons* or *neutral hadrons* depending on whether they are ECAL or HCAL clusters.
3. in this way all information in the detector is used to create a list of candidates which can be any of a muon, electron, photon, charged hadron or neutral hadron;
4. these are then used to construct composite detector objects such as jets if necessary.

Particle flow jets are constructed using the anti- $k_T$  algorithm[37]. This algorithm is both infrared and collinear (IRC) safe and preferentially clusters soft (low  $p_T$ ) jets with hard (high  $p_T$ ) jets to be robust in the LHC pileup conditions. These jets can then be additionally tagged as  $b$  or  $c$  (i.e. those containing a bottom or charm quark respectively) using techniques of the type described in[38]. There are also energy corrections applied to jets to account for pileup ( $\rho$  subtraction technique as described in Sec. 3.6), non-uniform detector response ( $p_T$  and  $\eta$  dependent corrections derived from MC simulation) and

data-MC differences (residual  $p_T$  and  $\eta$  corrections derived from  $\gamma$ +jet and  $Z$ +jet samples in data), see Ref.[39] for details. The clustering, and subsequent energy correction, of jets in this way also provides a measure of the missing transverse energy ( $\cancel{E}_T$ ), the amount of energy in an event taken away by undetectable particles such as neutrinos. A comparison of the jet energy resolution, as a function of jet  $p_T$ , for calorimeter jets and particle flow jets is shown in Fig. 3.8 for simulated data at  $\sqrt{s} = 7$  TeV.

A PF photon is quite different to the photons used in the analysis, whose reconstruction is described above in Sec. 3.3. PF is useful for tagging physics objects like  $\tau$  leptons or  $b$  quarks and for isolation sums (see following Sec. 3.5) but is not necessarily the best for reconstructing well measured objects like photons, electrons and muons.

### 3.5 Isolation

One way of differentiating between real (prompt) photons from Higgs decays and fakes is the use of isolation. One would expect that in the absence of pileup a photon from a Higgs is not in a jet and therefore would be isolated, which is to say there are no other particles (detector activity) in its vicinity. For a jet faking a photon (which is nearly always a  $\pi^0$ ) this is not the case and one would expect the  $\pi^0$  to be surrounded by additional hadronised particles (detector activity in the tracker, ECAL and HCAL). In the CMS  $H \rightarrow \gamma\gamma$  analysis isolation sums are used to distinguish prompt photons from fakes. Three variables are used which consider the isolation of each photon relative to activity in the surrounding environment. The procedure is to create a hollow cone around the photon candidate (of outer radius  $\Delta R_O$  and inner radius  $\Delta R_I$ , where  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ ) and sum the energy contained in that cone of PF candidates; charged hadrons, neutral hadrons and electrons/photons. These three variables are defined in the following way:

- **charged hadron isolation:** Sum of charged hadron PF candidates  $E_T$  in cone of  $\Delta R_O = 0.3$  and  $\Delta R_I = 0.02$ ;
- **neutral hadron isolation:** Sum of neutral hadron PF candidates  $E_T$  in cone of  $\Delta R_O = 0.3$  and  $\Delta R_I = 0.0$ ;
- **$e/\gamma$  isolation:** Sum of  $e/\gamma$  PF candidates  $E_T$  in cone of  $\Delta R_O = 0.3$  with a central  $\eta$  strip of 0.070 (0.015) removed for barrel (endcap) photons.

## 3.6 Pileup

There can be up to  $1.1 \times 10^{11}$  protons in each bunch at the LHC which results in multiple interactions (which can be as many as 50 primary vertices) per bunch crossing. This effect is known as pileup and can present a challenge when finding the primary vertex. This effect also produces additional energy in each event which originates from somewhere other than the primary vertex. To combat the latter of these two effects a technique called  $\rho$  subtraction is used to correct the energy of jets and isolation sums for pileup[40].  $\rho$  is defined as a per-event quantity and is computed by summing the energy in all the calorimeters and dividing by the calorimeter area and thus represents the median energy density in the detector per-event. The quantity  $\rho$  can then be used to subtract energy from isolation sums or alternatively its correlation with isolation sums in background samples can be exploited by multivariate analysis techniques.



# Chapter 4

## Common analysis components

This thesis describes three complementary analyses in the Higgs to two photons search at CMS. These differ in their photon selection, event selection, event classification (or categorisation) and statistical methods for extracting results. They are described in the following chapter (Chapter 5). However, there are many components which they share. These are detailed below.

As we have seen in Eq. 2.25 and 3.3, repeated below in Eq. 4.1 for convenience, the diphoton invariant mass is constructed from the two photon energies and the angle between them.

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos \alpha)} \quad (4.1)$$

Consequently, important considerations for this analysis are photon energy resolution and good opening angle resolution. The latter is completely dominated by the vertex resolution, as the position resolution of the photons (the location at which they hit the ECAL) is negligible in comparison. Details of how this is exploited in the analyses are given at the end of this chapter in Sections 4.3 and 4.4. The selection of events is described in the chapter after this (Chapter 5) alongside the categorisation, or binning, scheme whereby events which share similar signal to background ratios are collected into different classes of event, which take advantage of areas of phase space which share similar signal to background ratios. After a preliminary discussion of multivariate analysis techniques, the datasets, the triggering and the MC simulation are discussed.

## 4.1 Boosted Decision Trees

Multivariate analyses (MVA) are commonly used in High Energy Physics analyses to extract the maximum possible signal sensitivity in cases where the background rates are high. The advantage of Multivariate Analysis (MVA)s is that given a set of input variables a selection scheme can be built, to classify or correct events, in a multidimensional phase space to exploit differences between the signal and background in these variables and importantly in the correlations between them. A particular type of MVA which is used widely in this analysis is the Boosted Decision Tree (BDT). BDTs are preferred because they are more robust to the inclusion of variables which have little or no discriminating power. There are two broad types of BDT used, one is known as a regression BDT and the other as a classification BDT[41, 42, 43].

### 4.1.1 Classification BDT

A classification BDT will, given a set of input variables, assign a value (typically between  $-1$  and  $1$ ) to each event based on how signal-like that event is. This serves to collapse all the event information into one discriminating variable which can be used to classify differences between the signal and background. The input is provided as the probability distributions (which can be supplied as binned or unbinned data samples or as a functional form) of the background and signal for a set of “input variables”. The process involves construction of a series of Decision Trees (DT) complemented by a “boosting” step which serves to mitigate against “overtraining” on fluctuations within the training samples. This analysis chooses a particular type of decision tree boosting known as “gradient” boosting because it is more robust against outliers or mislabelled data points[44].

The Decision Tree (DT) is built by applying sequential cuts to the input variables and assessing the relative signal purity,  $p$ , in the sub-sample remaining after each cut

$$p = \frac{N_s}{N_s + N_b}, \quad (4.2)$$

where  $N_s$  and  $N_b$  are the sum of weights of the signal and background remaining in each sub-sample. A threshold criterion, known as the Gini index[44]  $p(1 - p)$ , is applied to decide whether to split the sample further. The process continues and the splitting is curtailed when either the threshold or the user defined maximum tree depth (number of subsamples allowed) is reached. The value of each cut is varied such that the signal purity,

$p$ , in each sub-sample is maximised. An event is assigned a value of  $-1$  or  $+1$  depending on whether it falls into a sub-sample with  $p > 0.5$  or not. Clearly some fraction of events will be misclassified where the actual number which get misclassified will depend on the discriminatory power available from the chosen input variables. In order to reduce this effect a series of DTs are trained and each assigned a weight derived by the “boosting” process.

We assign each DT as a member of a family of  $M$  functions,  $f(\vec{x}; \vec{a}_m)$ , which depend on the input variables,  $\vec{x}$ , and the set of cuts in that tree,  $\vec{a}_m$ . The object is to construct an overall decision tree which consists of the weighted average of each DT,

$$F(\vec{x}; \vec{\beta}, \vec{a}) = \sum_{m=0}^M \beta_m f(\vec{x}, \vec{a}_m) \quad \text{where } \vec{\beta} = (\beta_0, \beta_1 \dots \beta_M). \quad (4.3)$$

For the “adaptive” boost algorithm the input events for training the proceeding tree are reweighted by the fraction that get misclassified in the previous tree. That is, the background events get reweighted by the total fraction of background events which were classified as signal in the previous tree and the signal events get reweighted by the total fraction of signal events which were classified as background in the previous tree. In the “gradient” boosting procedure, which applies to the majority of BDTs used in this thesis, the weight for successive trees is obtained by minimising the deviation in the loss function (Eq. 4.4) each time a new tree is added<sup>1</sup>. The loss function, between the weighted tree response  $F(\vec{x}; \vec{\beta}, \vec{a})$  and the true output  $y$  obtained from the training sample, in this case is

$$L(F, y) = \ln(1 + e^{-2F(\vec{x})y}). \quad (4.4)$$

A common procedure when constructing a BDT to check for overtraining is to split both the background and signal into two independent samples. One is used to *train* the BDT and one is used to *test* the response of the output. Clearly one requires that both the training and independent test sample look the same in the output variable. This is usually quantified by use of a Kolmogorov-Smirnov test, which broadly speaking ascertains the probability that the training and test samples originate from the same underlying distribution[45].

---

<sup>1</sup> Friedman showed that for certain choices of loss function the “gradient” boost and “adaptive” boost procedures are identical, although this is not the case for a general loss function, nor the loss function of the form used in this thesis, shown in Eq. 4.4[43].

In this way the output of  $F(\vec{x}; \vec{\beta}, \vec{a})$  for a classification BDT will be a “semi-continuous” output<sup>2</sup> from  $-1$  to  $+1$  with signal events in general given a higher score than background events.

### 4.1.2 Regression BDT

A regression BDT is used to estimate the true value of some variable given the values and correlations of several other variables. They are commonly used for correcting the energy of a particular object, for example a photon. Given a MC source of photons the “true” energy is regressed from the position, shape and raw energy of the supercluster. For regression BDTs the output  $F(\vec{x}; \vec{\beta}, \vec{a})$  represents the estimated corrected energy and the boosting procedure targets minimising the deviation between this and the true energy in MC events.

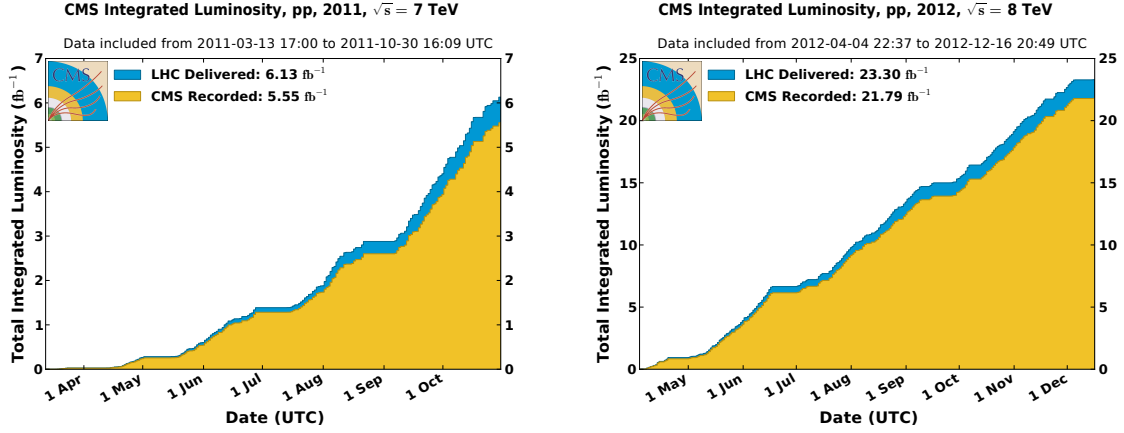
## 4.2 Data samples and triggering

The data consists of two independent samples of proton-proton collisions collected by the CMS experiment at the LHC in 2011 and 2012 with a centre-of-mass energy,  $\sqrt{s}$ , of 7 and 8 TeV, respectively. The total integrated luminosity of the two samples is  $5.1 \text{ fb}^{-1}$  and  $19.7 \text{ fb}^{-1}$  in 2011 and 2012, respectively, and they are collectively referred to as LHC Run 1. The response of the detector has changed considerably over this period and much of the variation is modelled by the MC simulation. Figure 4.1 shows the integrated luminosity delivered to and recorded by CMS during LHC Run 1.

Events are selected for the analysis by requiring they pass an asymmetric diphoton trigger with  $E_T$  thresholds of 26 (18) and 36 (22) GeV for the leading (trailing) photon in the 2011 and 2012 runs, respectively. The candidates are also required to have either a high value of  $R_9$  or to pass a loose calorimetric identification and isolation requirement. High trigger efficiency is achieved by selecting photon candidates which pass either requirement. The efficiency of the trigger for the analysis preselection is 99.5%.

---

<sup>2</sup> In the sense that each decision tree will give a discrete output of either signal-like or background-like such that the boosted output of the numerous trees in the forest contains several hundred (depending on the number of trees) discrete values in the range  $[-1, 1]$ .



**Figure 4.1:** The total integrated luminosity delivered to and recorded by CMS during the 2011 (left) and 2012 (right) run periods. Due to down time of various subsystems in CMS during run periods, particularly the ECAL, the recorded luminosity given here is not exactly equivalent to the integrated luminosity of the datasets used in the analysis[46, 47].

### 4.2.1 Monte Carlo Simulation

Accurate simulation of detector effects and efficiency is highly important. Knowledge of the expected Higgs signal shape is clearly essential and although the size and shape of the  $m_{\gamma\gamma}$  background is entirely data driven when extracting results, simulating the kinematics, shower shape and resolution properties of the background is important when training the selection and optimising the categorisation of events.

As explained in Chapter 2 the two main production mechanisms for a SM Higgs boson at the LHC are  $ggH$  and  $VBF$ . Typically the latter is produced at much higher Higgs  $p_T$  and this feature is exploited in the analysis (see Fig. 2.2). Consequently, it is important to model the  $p_T$  distribution of these two production modes accurately. The signal samples for these two processes are generated using POWHEG[48, 49] at NLO interfaced with PYTHIA[50] including a reweighting factor which matches their  $p_T$  spectrum to that when including the NNLO and NNLL terms. For the associated production modes, with a  $W^\pm$ ,  $Z$  or  $t$  quarks, ( $VH$  and  $t\bar{t}H$ ) only PYTHIA is used. The SM Higgs boson cross sections and branching fractions, and their uncertainties, are taken from Ref[25]

The spin-2 graviton with minimal couplings,  $2_m^+$ , has two production mechanisms, one via gluon-fusion ( $ggX$ ) and one via quark-antiquark annihilation ( $q\bar{q}X$ ). The graviton samples are generated using the JHU generator[51]. In these samples the signal events are reweighted such that the graviton  $p_T$  spectrum matches the Higgs  $p_T$  spectrum in

the SM signal samples. The kinematic properties of a spin 2 *graviton-like* Higgs are not well defined and can depend on the specifics of the model in question. Matching the  $p_T$  spectrum of the spin-2 samples with the SM spin-0 ensures that no discrimination arising from model dependence of the  $p_T$  distribution will arise.

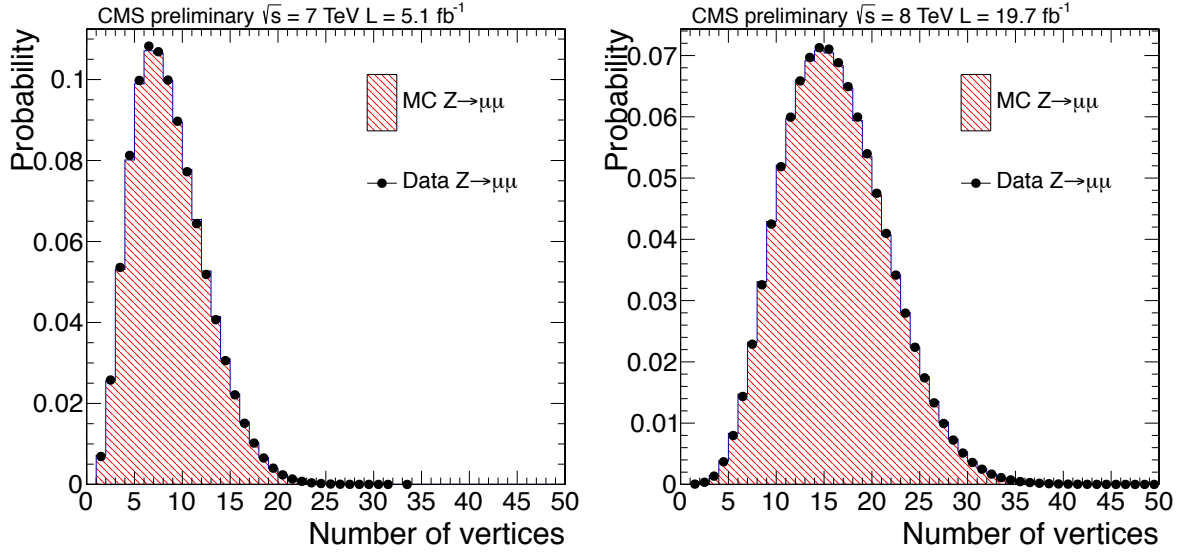
The simulated background samples are used solely for cut and category optimisations and training of multivariate discriminants. The background which contains the QCD continuum of prompt photons (referring back to Chapter 2 these are produced by Born and box diagrams) is simulated using SHERPA[52] at 8 TeV and MADGRAPH[53] at 7 TeV. The prompt-fake and fake-fake backgrounds, in which one or both photons are faked by a neutral meson (usually a  $\pi^0$ ) reconstructed as a photon, are generated using PYTHIA. Samples of  $Z \rightarrow e^+e^-$ ,  $Z \rightarrow \mu^+\mu^-$  and  $Z \rightarrow \mu^+\mu^-\gamma$  used for data/MC comparisons are generated with POWHEG.

All of these *generator level* samples are then run through the full CMS detector simulation using GEANT4[54]. This includes the effect of overlapping vertices (pileup) and detector effects (such as noise and crystal degradation) in four time periods; Run2011 ( $5.1 \text{ fb}^{-1}$ ), Run2012AB ( $5.3 \text{ fb}^{-1}$ ), Run2012C ( $7.1 \text{ fb}^{-1}$ ) and Run 2012D ( $7.3 \text{ fb}^{-1}$ ).

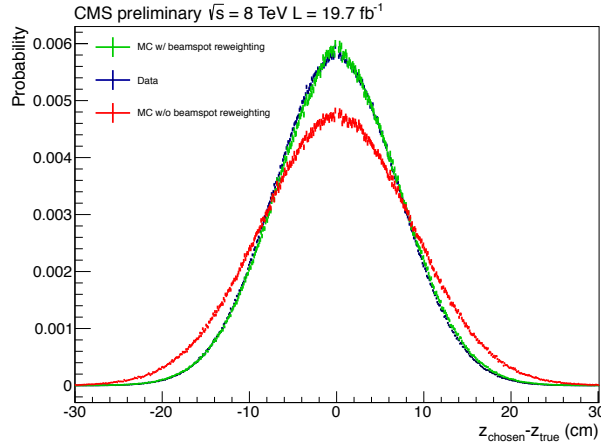
### 4.2.2 Pileup and beamspot reweighting

An important difference between the simulated samples and the data which can have a large impact on the analysis is the distribution of the number of primary vertices. The *pileup* in the event affects many important analysis variables, for example photon shower shape and photon isolation as well as the diphoton invariant mass if the chosen vertex is wrong. Consequently the MC events are reweighted such that the pileup distribution matches that in data. The reweighting technique is validated using  $Z \rightarrow \mu^+\mu^-$  events as shown in Fig. 4.2 for the 7 and 8 TeV samples.

When the chosen vertex is incorrect the mass resolution is dominated by the spread in position of the pileup vertices (known as the beamspot width,  $\sigma_z^{beamspot}$ ). Accurate modelling of this spread is important so that the resolution of wrong vertex events in simulation matches that in data. The MC sample overestimates the beamspot spread by some 20% so a simple reweighting is implemented for MC events in which the wrong vertex is chosen (as the effect is negligible for events in which the chosen vertex is correct) such that the distribution of the distance between the chosen vertex position and the



**Figure 4.2:** Distribution of the number of reconstructed vertices in the 2011 (left) and 2012 (right) run periods. Calculated using the Deterministic Annealing algorithm in [55] for  $Z \rightarrow \mu^+ \mu^-$  events in data (black dots) and MC events (red histogram) after reweighting.



**Figure 4.3:** Distribution of  $\Delta z$  (the distance between the chosen vertex and the true vertex in the  $z$  direction) for data (black), the MC simulation (red) and the MC simulation after beam spot reweighting (green) for  $Z \rightarrow \mu^+ \mu^-$  events.

true vertex position,  $\delta z = z_{\text{chosen}} - z_{\text{true}}$ , match between data and MC. The effect with and without reweighting compared to data is shown in Fig. 4.3.

### 4.3 Energy measurement of photons

The photon energy obtained from the supercluster sum described in Section 3.2.2, even when including the intercalibration and transparency corrections shown in Fig. 3.7, does not give the most optimal resolution for the energy measurement of photons at CMS. On top of this energy (known as the *raw* supercluster energy,  $E_{raw}$ ) it is also valuable to correct for additional energy losses. These arise from bremsstrahlung; where the photon converts in the material upstream of the ECAL and the two electrons radiate additional photons and thus some of the photon shower is missed, and from local non-containment of the shower; where some energy is lost due to small gaps between ECAL crystals and larger gaps between “modules” or sections of crystals. These corrections are obtained using a specialised regression BDT (see Sec. 4.1) trained on a MC source of prompt photons from a sample containing photons and jets. The BDT targets accurate measurements of individual photons’ energies by correcting the raw supercluster energy and provides an estimate for the energy resolution of each photon given the position and shower shape of the supercluster. The training is done separately for barrel and endcap photons (as the cluster shapes look very different for these two distinct regions) and is also performed separately for the 7 and 8 TeV datasets. The following input variables are used:

- the global position of the supercluster in  $\eta$  and  $\phi$ ;
- a collection of shower shape variables which aim at providing information on the likelihood and location of a photon conversion and the degree of showering in the material:
  - the  $R_9$  of the supercluster (as previously described in Section 3.2.2);
  - the ratio of the  $5 \times 5$  crystal energy to the raw supercluster energy (equivalent to  $R_{25}$ );
  - the energy weighted  $\eta$ -width and  $\phi$ -width of the supercluster (in other words the spread of the shower);
  - the number of basic clusters in the supercluster;
  - the ratio of energy in the HCAL behind the supercluster to the ECAL energy of the supercluster,  $H/E$ ;
  - the ratio of the preshower energy to the raw supercluster energy (endcap only).



- a collection of the seed crystal and the seed cluster variables which aim at providing information about energy lost through gaps and cracks between crystal and crystal modules:
  - the relative energy and position of the seed cluster;
  - the local energy covariance matrix;
  - energy ratios between the seed and the  $3 \times 3$  and  $5 \times 5$  areas around the seed;
  - the  $\eta$  and  $\phi$  index of the seed crystal and the position of the seed cluster relative to the crystal centre.
- additionally, the number of primary vertices and the median energy density,  $\rho$ , (see Sec. 3.6) are included to account for residual energy scale effects from pileup.

The regression is trained using an additional feature to that described in Sec. 4.1 whereby the target is to predict the full probability distribution of the ratio of the true energy to the raw energy,  $E_{true}/E_{raw}$ . The target is a double Crystal Ball[56] distribution which consists of a Gaussian core and power law tails on either side. This can be fully parametrised by six variables, the Gaussian mean and width ( $\mu$ ,  $\sigma$ ), the power parameters ( $n_L$ ,  $n_R$ ) and the power law tail cutoff parameters ( $\alpha_L$ ,  $\alpha_R$ ). Each of these parameters has a non-parametric dependence on the input variables,  $\vec{x}$ , and this is *learned* by the regression training whilst simultaneously minimising the likelihood,

$$-\ln \mathcal{L} = - \sum_{MC \text{ photons}} \ln p[E_{true}/E_{raw} | \mu(\vec{x}), \sigma(\vec{x}), \alpha_L(\vec{x}), \alpha_R(\vec{x}), n_L(\vec{x}), n_R(\vec{x})], \quad (4.5)$$

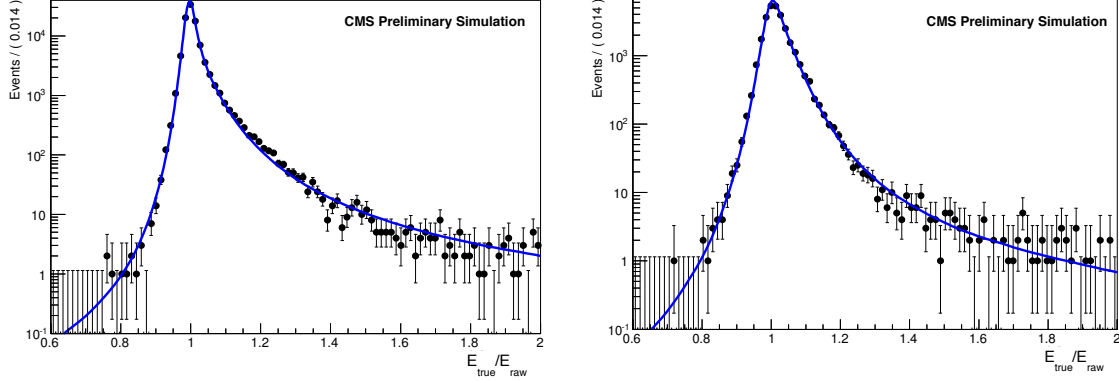
for the double Crystal Ball distribution,  $p$ . The most probable value for the true energy estimate of each photon is then given by,

$$E(\vec{x}, E_{raw}) = \mu(\vec{x}) E_{raw} \quad (4.6)$$

and the per-photon energy resolution is given by,

$$\frac{\sigma_E(\vec{x}, E_{raw})}{E(\vec{x}, E_{raw})} = \frac{\sigma(\vec{x})}{\mu(\vec{x})}. \quad (4.7)$$

In this way the regression predicts the full probability distribution of  $E_{true}/E_{raw}$  per photon given a particular configuration of the input variables,  $\vec{x}$ , and provides an estimate of the optimal energy correction and the energy resolution per photon. A comparison



**Figure 4.4:** A comparison of the predicted probability density of  $E_{true}/E_{raw}$  from the regression training (blue line) to the distribution in a statistically independent MC sample (black points) for barrel photons (left) and endcap photons (right).

of this distribution with a statistically independent MC sample is shown for the 8 TeV training in Fig. 4.4.

### 4.3.1 Correcting for residual discrepancies between data and Monte Carlo simulation

After application of the energy regression correction there are some remaining discrepancies between data and MC simulation. These residual effects are accounted for using  $Z \rightarrow e^+e^-$  events in data and simulation to correct the energy scale in the data and to apply an additional smearing term to the MC events with systematic uncertainties propagated through the analysis to account for the uncertainties on these corrections.

#### Energy scale corrections to the data

The supercluster energy is identical for electrons and photons so by correcting the supercluster energy scale to a known source, namely the mass of the  $Z$ -boson, in dielectron decays the smaller residual energy scale effects are accounted for. When reconstructing the dielectron decays to derive these corrections, electrons are reconstructed as photons and the  $Z$  mass calculated in the same way as for the diphoton invariant mass in Eq. 3.3, where the electron energy is obtained from the supercluster alone and the dielectron opening angle is obtained from the tracks. This can be done several times to account for various different effects. In the first stage scale corrections are derived in bins of

time (run range) and  $\eta$ . After applying these corrections, further, much smaller, residual effects are accounted for in bins of  $R_9$  (the size of the effect is different for converted and non-converted photons). After applying both of these a further step is taken for the 8 TeV data in the barrel to derive residual corrections in bins of  $E_T$ . Consequently the total scale correction is a product of three corrections in 59 bins of time  $\times$  4 bins in  $\eta$   $\times$  2 bins in  $R_9$   $\times$  6 bins in  $E_T$  (where the last is applied for the 8 TeV barrel photons only).

The strategy for deriving these corrections is to take  $Z \rightarrow e^+e^-$  events in data and MC simulation and extract the invariant dielectron mass in the relevant bin of interest. This mass distribution is fitted with a convolution of a relativistic Breit-Wigner (designed to handle the underlying Z line shape[4]) and a Crystal Ball function which models the calorimeter resolution effects and bremsstrahlung losses in the material upstream of the ECAL. The Breit-Wigner parameters are fixed to the PDG values of  $M_Z = 91.188$  GeV and  $\Gamma_Z = 2.495$  GeV[4] whilst the Crystal Ball parameters which model the detector effects are allowed to float. The scale correction,  $\Delta E$ , is then defined as the relative difference between the Crystal Ball peak in data and simulation,

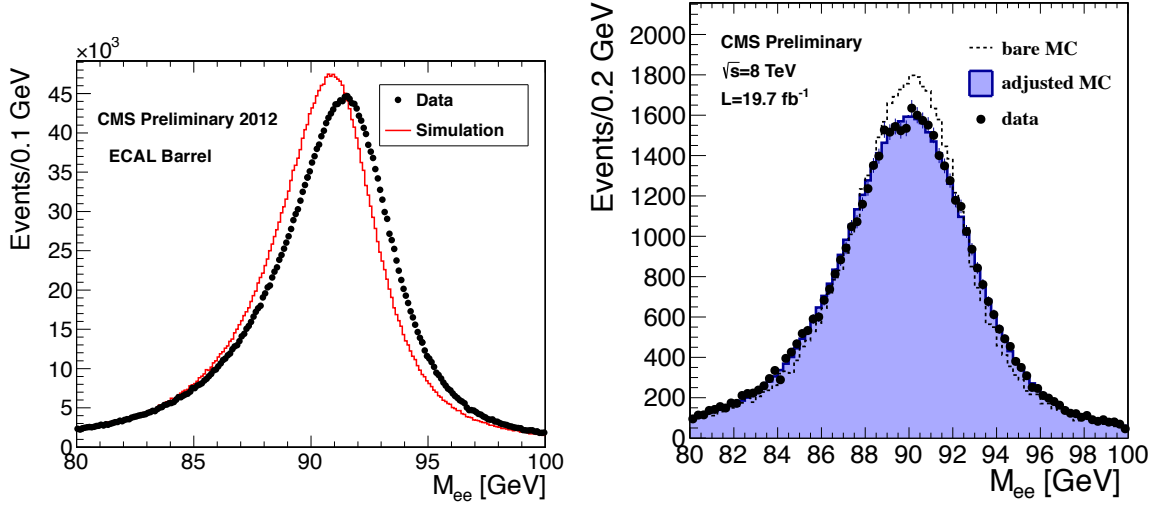
$$\Delta E = \frac{m_{data} - m_{MC}}{M_Z}. \quad (4.8)$$

### Energy resolution smearing for the Monte Carlo events

A similar method is used to extract a smearing factor that can be applied to the MC events such that the width of the invariant mass distribution in  $Z \rightarrow e^+e^-$  decays matches between data and MC events. This is done in 4 bins of  $\eta$   $\times$  2 bins of  $R_9$  and is parametrised as the quadratic sum of two resolution components: a constant term,  $\Delta C$ , and a stochastic term,  $\Delta S$ , which aims to model the expected resolution effects explained in Eq. 3.5 in Sec. 3.2.2. The smearing term,  $\Delta\sigma$ , is parametrised as,

$$\Delta\sigma = \frac{\Delta S}{\sqrt{E_T}} \oplus \Delta C. \quad (4.9)$$

The effect of the scale and smearing corrections is shown for the  $Z \rightarrow e^+e^-$  data and MC samples in Fig. 4.5. Figure 4.6 shows the  $Z \rightarrow e^+e^-$  invariant mass distribution for data and MC events at 8 TeV for events passing the analysis preselection. The discrepancy between the data and the MC simulation in the tails of the dielectron invariant mass distribution, shown in the right hand plot of Fig. 4.6, is considerably

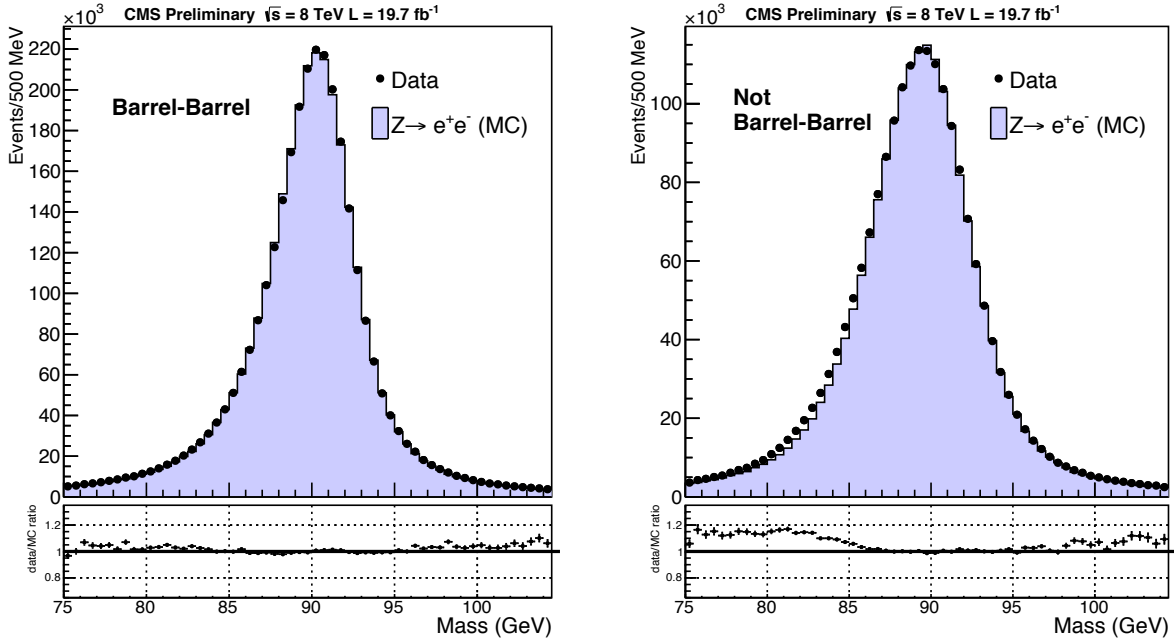


**Figure 4.5:** The  $Z \rightarrow e^+e^-$  invariant mass shape comparison of data and MC events before (left) and after (right) the scale and smearing corrections are applied. Shown for 8 TeV electrons (reconstructed as photons) in the barrel (left) and in the barrel with  $R_9 \geq 0.94$  (right).

reduced (from  $\sim 20\%$  to  $\sim 10\%$ ) after the full analysis event selection is made (not just the preselection shown here). Furthermore, the analysis sensitivity is decreased by less than 5% when events with at least one photon in the endcap are removed and consequently the remaining discrepancy is not considered as significant. Each of the scale and resolution corrections has an associated uncertainty and these uncertainties are propagated per photon through the analysis. There are also additional uncertainties included which account for differences between electrons and photons and the difference between the  $Z$  mass scale (around 90 GeV) and the Higgs mass scale (around 125 GeV). Systematic uncertainties are described in more detail in Sec. 6.3.

## 4.4 Vertex reconstruction

The resolution on the opening angle has a negligible effect if the correct vertex can be found within 10 mm of the true interaction point. As seen in Sec. 4.2.2 the beamspot has an RMS spread of about 5 cm in the  $z$  direction and there is an average of  $\sim 20$  vertices per bunch crossing. Because the beam direction is along the  $z$ -axis the spread of the vertex in the  $x$  and  $y$  directions is tiny ( $< 0.5$  mm) and consequently mismeasurement of the primary vertex in the  $x$ - $y$  plane is small and has no impact on the mass resolution. By assigning the correct vertex to the diphoton pair, using other information in the tracking

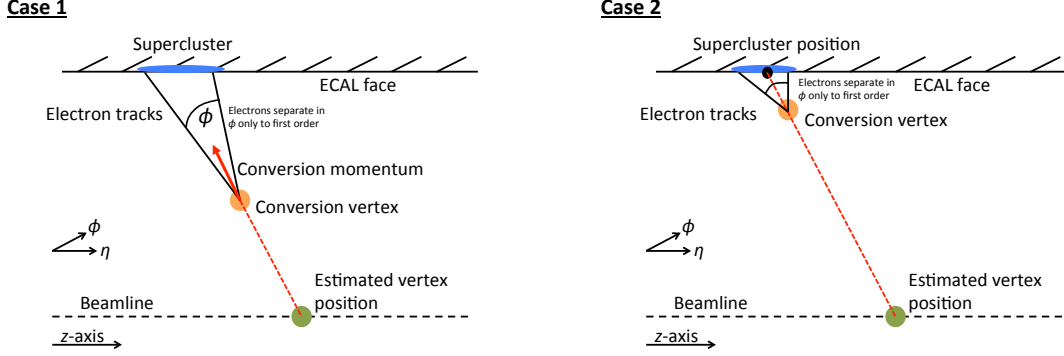


**Figure 4.6:** The  $Z \rightarrow e^+e^-$  invariant mass distribution at 8 TeV for electrons reconstructed as photons when both electrons are in the barrel (left) and when at least one is not in the barrel (right). Shown for data (black points) and MC events (blue histogram) which pass the analysis preselection when the electron veto is inverted.

system, most of the mass resolution can be preserved. The method used to extract the primary vertex is a classification BDT which exploits the correlation between the diphoton pair and the recoiling tracks from the underlying interaction as well as additional information in the tracking system if there is a photon conversion pair. The output of this per-vertex BDT is evaluated for each vertex in the event and the primary vertex is assigned as the one with the highest value of the BDT output (i.e. the value nearest 1.). In addition, it is possible to construct another BDT whose output is proportional to the probability that the chosen vertex is the correct one (described in Sec. 4.4.1). This probability becomes a useful discriminating variable for the analyses later on.

The vertex BDT uses the following input variables:

- $\sum_i |\vec{p}_T^i|^2$  - the sum of the transverse momentum squared of all of the tracks which originate from this vertex, representing how hard the interaction is at this vertex.
- $\frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|} \cdot \sum_i \vec{p}_T^i$  - the dot product between the transverse momentum of the diphoton system and the sum of all other tracks originating from this vertex, representing the recoil of the tracks relative to the diphoton system.



**Figure 4.7:** A representation of the two methods for locating the primary vertex using photon conversion information. The left plot is for cases where the conversion occurs early enough in the tracker that the two electron tracks can be used to construct the converted pair momentum, which is then combined with the conversion vertex position to point back to the beam line. The right plot is for cases where the conversion occurs late in the tracker and the energy weighted supercluster position and the conversion vertex position are used to point back to the beam line.

- $(|\sum_i \vec{p}_T^i| - |\vec{p}_T^\gamma|) / (|\sum_i \vec{p}_T^i| + |\vec{p}_T^\gamma|)$  - the asymmetry between the diphoton system and the other tracks originating from this vertex.
- $|z_v - z_c| / \sigma_c$  - this is added for events which contain at least one photon conversion where  $z_v$  is the  $z$  position of the vertex in question and  $z_c$  and  $\sigma_c$  are the estimated  $z$  position of the vertex from conversion information and its approximate error as defined below.

For events which contain at least one photon conversion, the conversion tracks and/or the conversion momentum can be used to point back to the beam line and estimate the vertex position. This can be achieved in one of two ways. In cases where the conversion occurs early, i.e. in one of the first layers of the tracking system, then the electron pair from the conversion will leave two clean and distinct tracks. This means that the momentum of the conversion pair can be accurately reconstructed and used to point from the conversion vertex position back to the beam line and thus the nearest primary vertex. In cases where the conversion occurs late in the tracking system, there are not enough track hits to accurately reconstruct the momentum of the conversion pair. However the incident position of the photon at the ECAL face is well known in this case, so the line which connects the ECAL position with the conversion vertex can be used to point back to the beam line. This is diagrammatically represented in Fig. 4.7 for both cases.

For Case 1 conversions the primary vertex  $z$  position is calculated as,

$$z_c = z_{conv} - r_{conv} \cot(\alpha), \quad (4.10)$$

where  $z_{conv}$  is the  $z$  position of the conversion vertex,  $r_{conv}$  is the distance of the conversion vertex from the beam line and  $\alpha$  is the angle between the beam line and the conversion momentum.

For Case 2 conversions the primary vertex  $z$  position is calculated as,

$$z_c = \frac{r_{conv} z_{SC} - r_{SC} z_{conv}}{r_{conv} - r_{SC}}, \quad (4.11)$$

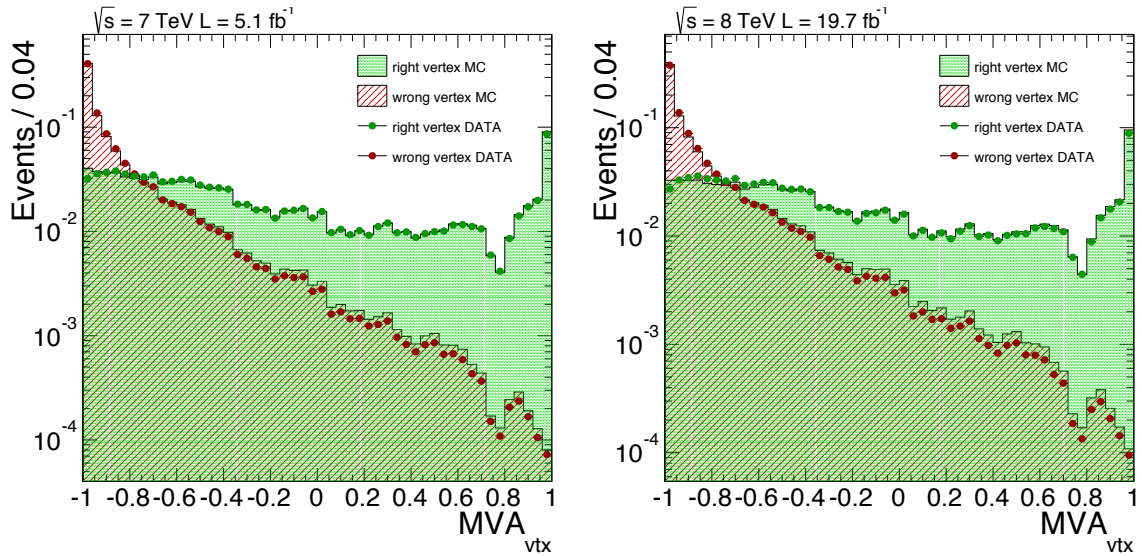
where  $z_{conv}$  and  $z_{SC}$  are the  $z$  positions of the conversion vertex and supercluster respectively, and  $r_{conv}$  and  $r_{SC}$  are the distance of the conversion vertex and the supercluster from the beamline.

There are six regions of the tracking system (refer back to Fig. 3.2). When the conversion vertex is located in one of the inner regions; Pixel Barrel, Pixel Forward or TID, the Case 1 conversion information is included in the BDT, otherwise the Case 2 conversion information is used. The resolution on the primary vertex position in conversions is estimated per tracking region using  $\gamma$ +jet events in data for which the primary vertex efficiency is high and the photon converts. Using these events, the conversion resolution,  $\sigma_c$ , is calculated as the effective width<sup>3</sup> of the distribution of the difference,  $\Delta z = z_v - z_c$ , between the estimated  $z$  position of the primary vertex without any conversion information,  $z_v$ , and the estimated  $z$  position of the primary vertex when using conversion information alone,  $z_c$ . Consequently the fourth input variable to the BDT, shown in the list above as  $|z_v - z_c|/\sigma_c$ , is effectively a pull distribution for the conversion vertex. The BDT will favour vertices whose value of this variable is near zero.

The BDT is trained on a sample of  $H \rightarrow \gamma\gamma$  MC events. It is tested with a statistically independent sample and further validated using  $Z \rightarrow \mu^+\mu^-$  decays in data and MC samples. The efficiency is measured in data using the  $Z \rightarrow \mu^+\mu^-$  channel where the muon tracks are removed from the BDT variables to simulate a diphoton-like situation in data. The BDT response is shown for  $Z \rightarrow \mu^+\mu^-$  data and MC events for both the signal (right vertex) and background (wrong vertex) in Fig. 4.8. The chosen primary vertex is the one which gives the highest score BDT output. The efficiency of the vertex

---

<sup>3</sup>Half the narrowest interval which contains 68.3% of the distribution



**Figure 4.8:** The vertex BDT response for  $Z \rightarrow \mu^+\mu^-$  events in data (points) and MC events (filled histogram) for the primary vertex (green) and the background pileup vertices (red).

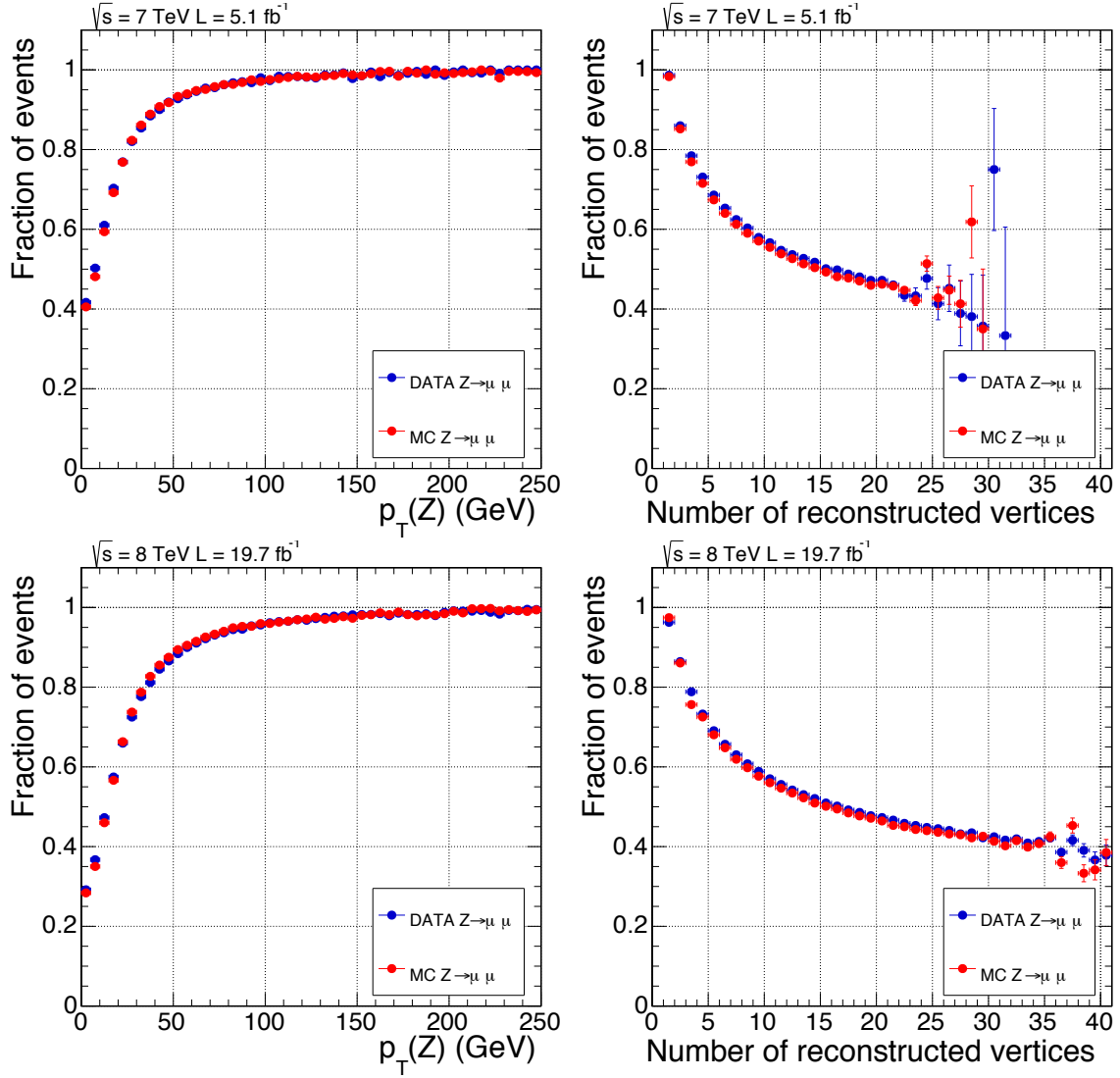
selection as a function of the  $Z$   $p_T$  and the number of reconstructed vertices as measured in  $Z \rightarrow \mu^+\mu^-$  data and MC samples is shown in Fig. 4.9.

#### 4.4.1 Estimating the per-event probability that the correct vertex is chosen

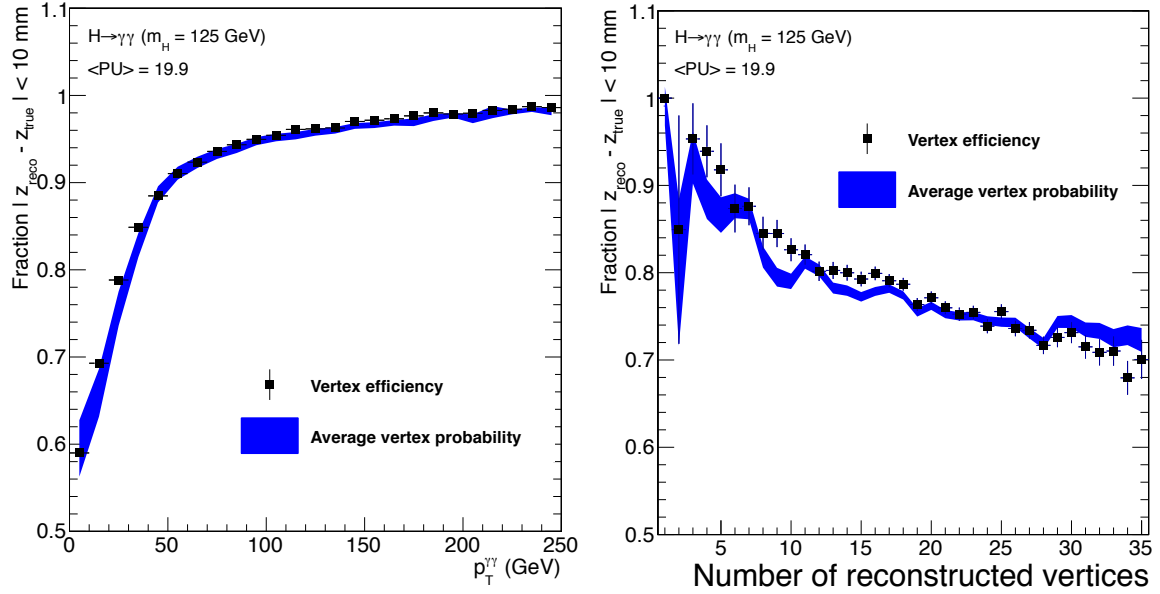
The total efficiency for assigning the correct vertex using the method described in the preceding section is at the level of 75% during 2012 running conditions, where the correct vertex is defined as being within 10 mm of the true vertex. This means that for around 25% of preselected events the mass resolution is dominated by the vertex resolution. Consequently, it is important to ascertain the probability that the chosen vertex is the correct one. An additional specific BDT is constructed to address exactly this topic. The input variables used for this BDT are:

- the  $p_T$  of the diphoton system;
- the number of vertices in each event;
- the value of the per-vertex BDT described above, for the three vertices with the highest score;





**Figure 4.9:** The chosen vertex efficiency as measured in  $Z \rightarrow \mu^+ \mu^-$  data and MC simulation as a function of  $Z$   $p_T$  (left) and number of reconstructed vertices (right) for the 7 TeV (top row) and 8 TeV (bottom row) data samples.



**Figure 4.10:** A comparison of the true vertex efficiency (black points) and the average vertex probability (blue band) for a statistically independent MC Higgs sample simulated with 2012 running conditions.

- the  $z$  distance,  $\Delta z$ , between the chosen vertex and the second and third choice vertices;
- the number of photon conversions used; either 0, 1 or 2.

There is a linear relation between the response of this BDT and the correct vertex efficiency (or probability). This is used to analytically obtain the per-event correct vertex probability for a given event. Figure 4.10 shows that this estimation reproduces the required vertex efficiency as a function of Higgs  $p_T$  and number of reconstructed vertices.

## 4.5 Event preselection

A simple and loose preselection is applied to all photons before they enter the analysis. The preselection requirements are identical for all analysis approaches and are designed to remove some *fake* photons whilst maintaining near 100% trigger efficiency. The variables used for preselection are defined below and the preselection cuts are described in Table 4.1.

- $H/E$  - The ratio of hadronic energy in the HCAL tower behind the supercluster to the electromagnetic energy in the supercluster. Neutral jets which fake photons

typically leave a fraction of their energy in the HCAL so there is a requirement that the value of this variable is small.

- $\sigma_{ini\eta}^2$  - The RMS spread of the shower in the  $\eta$  direction. Multiple showers of a  $\pi^0$ , or more than one  $\pi^0$ , result in a wider shower in  $\eta$  (as the  $\pi^0$  decay product photons are separated in space). This cannot be exploited in the  $\phi$  direction because conversion electrons get separated by the magnetic field, however single photons, even when converted, occupy a narrow region in  $\eta$ . The separation of the two photons from a  $\pi^0$  is minimal when they share the energy equally and given that typically  $p_T \gg m_\pi$  the separation is close to minimal for most  $\pi^0$  decays. Taking the transverse plane in the barrel, the separation  $d = 2Rm_\pi/p_T$  (where  $R$  is the radius of the barrel), which for  $p_T = 40$  GeV gives a value of  $d = 8$  mm  $\sim 0.006$  in  $\eta$ . By referring to Table 4.1 it is clear that the preselection requirement is quite loose.
- $ISO_{ECAL}$  - The total  $\rho$ -corrected electromagnetic energy in a cone of radius 0.4 in  $(\eta, \phi)$  around the photon candidate - see Sec. 3.5.
- $ISO_{HCAL}$  - The total  $\rho$ -corrected hadronic energy in a cone of radius 0.4 in  $(\eta, \phi)$  around the photon candidate - see Sec. 3.5.
- $ISO_{Tracks}$  - The total  $\rho$ -corrected track energy in a cone of radius 0.4 in  $(\eta, \phi)$  around the photon candidate - see Sec. 3.5.
- $ISO_{PFCh}$  - The total  $\rho$ -corrected particle flow charged hadron energy in a cone of radius 0.4 in  $(\eta, \phi)$  around the photon candidate - see Sec. 3.5.

In addition to the above an electron veto is applied to prevent contamination of the photon sample with electrons which originate from Drell-Yan interactions. This is achieved by removing photon candidates whose supercluster is matched to an electron track which has no missing hits in the innermost tracking region.

## 4.6 Using $Z$ decays for validation and efficiency measurements

Whilst no known “standard candles” with high statistics exist for high  $p_T$  photons in the LHC environment a powerful control source for the  $H \rightarrow \gamma\gamma$  decay in both data and MC simulation is the  $Z \rightarrow e^+e^-$  decay. From an ECAL interaction view point

**Table 4.1:** Preselection cut values.

	Barrel		Endcap	
$R_9$	$H/E$	$\sigma_{i\eta i\eta}^2$	$H/E$	$\sigma_{i\eta i\eta}^2$
$\leq 0.9$	$< 0.075$	$< 0.014$	$< 0.075$	$< 0.034$
$> 0.9$	$< 0.082$	$< 0.014$	$< 0.075$	$< 0.034$
	Both Barrel and Endcap			
$R_9$	$ISO_{ECAL}$	$ISO_{HCAL}$	$ISO_{Tracks}$	$ISO_{PFCh}$
$\leq 0.9$	$< 4 \text{ GeV}$	$< 4 \text{ GeV}$	$< 4 \text{ GeV}$	$< 4 \text{ GeV}$
$> 0.9$	$< 50 \text{ GeV}$	$< 50 \text{ GeV}$	$< 50 \text{ GeV}$	$< 4 \text{ GeV}$

electrons are very similar to photons and the  $Z$  is relatively near the relevant Higgs search range in mass. The differences between the  $Z$  and the Higgs, in both their mass and  $p_T$  distribution, and also the differences between electrons originating from a  $Z$  and photons originating from a Higgs are important systematic uncertainties on the Higgs mass scale and resolution. By inverting the electron veto usually applied in the preselection, the Higgs to two photon analysis can be identically replicated but with the very pure diphoton sample replaced with a pure dielectron sample. One additional process which can be used as a direct control for photons is the  $Z \rightarrow \mu^+\mu^-\gamma$  decay although the statistics, even with the LHC luminosity, are very low. Many of the input variables used in training the BDTs and cuts are validated with both  $Z \rightarrow e^+e^-$  and  $Z \rightarrow \mu^+\mu^-\gamma$  data/MC comparison plots. An example has been shown in a previous figure (see Fig. 4.6) for the reconstructed dielectron mass for events passing the preselection described in Table 4.1.

As previously shown (in Sec. 4.3.1) data/MC comparisons of the  $Z \rightarrow e^+e^-$  decay are used to derive scale corrections for the data and smearing of the MC simulation. Discrepancies between data and MC in  $Z \rightarrow e^+e^-$  decays of important analysis variables are accounted for by introducing systematic uncertainties to cover them. In addition the “tag and probe” method [57] is used on  $Z \rightarrow e^+e^-$  decays to evaluate the signal efficiency for the preselection and analysis cuts. Several stages of the analysis are validated in this way and where appropriate systematic uncertainties are included to account for any data/MC differences. Although the numbers and uncertainties themselves are derived from  $Z \rightarrow e^+e^-$  samples (because of the much higher statistics), they are always cross-checked with the  $Z \rightarrow \mu^+\mu^-\gamma$  sample.

# Chapter 5

## Selection and Categorisation

This thesis presents results of three different analyses used in the Higgs to two photon search at CMS. The nominal results and properties are obtained from the so-called Mass-Factorised MVA (MFM) analysis, which uses multivariate methods optimised specifically to search for a SM Higgs boson to select and categorise events. This has a fully parametric definition of the diphoton invariant mass spectrum where the signal shape is derived from MC simulation and the background shape from data. Events are selected by requiring that they pass a cut on the output of a BDT trained to distinguish photons from jets (the *photon ID BDT*) and furthermore that they pass a cut on an event-level classifier (the *diphoton BDT*) designed to collapse photon kinematics, mass resolution and photon quality into a single discriminating variable. Using the output of this event-level classifier a number of analysis categories are defined to provide the optimal search sensitivity for a SM Higgs boson.

The second analysis is the Sideband MVA (SMVA) analysis which is a cut and count method and serves to cross-check the background estimation in the MFM analysis, the most significant unknown in a search like this, by extracting the background under the signal region from sidebands in the diphoton invariant mass spectrum. The SMVA uses the same event selection as the MFM although the categorisation is done differently.

The third analysis is the Cuts in Categories (CiC) analysis which is designed for simplicity and robustness as a cut based approach and, owing to its low level of model dependence, is used for statistical tests which attempt to ascertain the spin of the observed boson.

## 5.1 Event selection

There are two complementary event selections used in the three analyses. The CiC analysis uses a cut-based photon selection whilst the MFM and SMVA analyses use a series of BDTs to first select photons and then select events. The ultimate aim is a selection which accepts two prompt photons (i.e. does not contain any fakes) and exploits regions of phase space which have a narrow mass resolution and high signal to background ratio.

All events must contain at least two photons which pass  $p_T^{\gamma_1}/m_{\gamma\gamma} > 1/3$  (for the leading photon) and  $p_T^{\gamma_2}/m_{\gamma\gamma} > 1/4$  (for the subleading photon) and the invariant mass of the diphoton pair must be in the range  $100 < m_{\gamma\gamma} < 180$  GeV. The reason for choosing  $p_T$  cuts which slide with  $m_{\gamma\gamma}$  is to avoid any Higgs mass,  $m_H$ , dependence in the selection. In the case where there are more than two photon candidates in an event, the two photons with the highest  $p_T$  are chosen.

### 5.1.1 Selection using cuts in categories

The selection cuts are optimised for photons in four non-overlapping categories to take advantage of the different photon energy resolution between the barrel and the endcap and between converted and non-converted photons. The cuts, applied to the leading and sub-leading photons, are used to select diphoton events. The categories are defined as  $|\eta| < 1.444$  or  $|\eta| > 1.556$  (i.e. either barrel or endcap) and  $R_9 < 0.94$  or  $R_9 \geq 0.94$  (i.e. either converted or non-converted) and the values for each of the cut variables (shown in Table 5.1) are chosen by targeting a specific signal to background ratio ( $S/B$ ) of 0.05. The procedure to define the cuts is as follows:

- a set of loose cuts are defined as the starting values;
- the “N-1” distribution of each cut variable is produced. This is the distribution of each variable after the cuts on all other variables have been applied;
- a smooth curve is fitted to the distribution of the  $S/B$  ratio versus the cut variable;
- the cut value is chosen by evaluating the curve for the required  $S/B$  value and thus requires that photon candidates come from a region of phase space which has at least this  $S/B$  ratio or higher;

- do the same for all the other variables.

Consequently a stable set of cut values is obtained by iterating this procedure a few times. Each cut will then select events with the same purity ( $S/B$ ) and thus the efficiency for selected photons is maximised for the chosen purity level. As a result the cuts in the endcap are much tighter than in the barrel, similarly the cuts for low  $R_9$  photons are much tighter than the cuts for high  $R_9$  photons. The cut variables are described below and the cut values shown in Table 5.1. The cut setting procedure is optimised on a signal sample of  $H \rightarrow \gamma\gamma$  MC events (with  $m_H = 120$  GeV) and a background sample of  $\gamma$ +jet events. The photon efficiency of the cuts for the four different classes of photon is shown in Fig. 5.1 as a function of the supercluster position,  $\eta$ , and the photon  $p_T$ . The cut variables are:

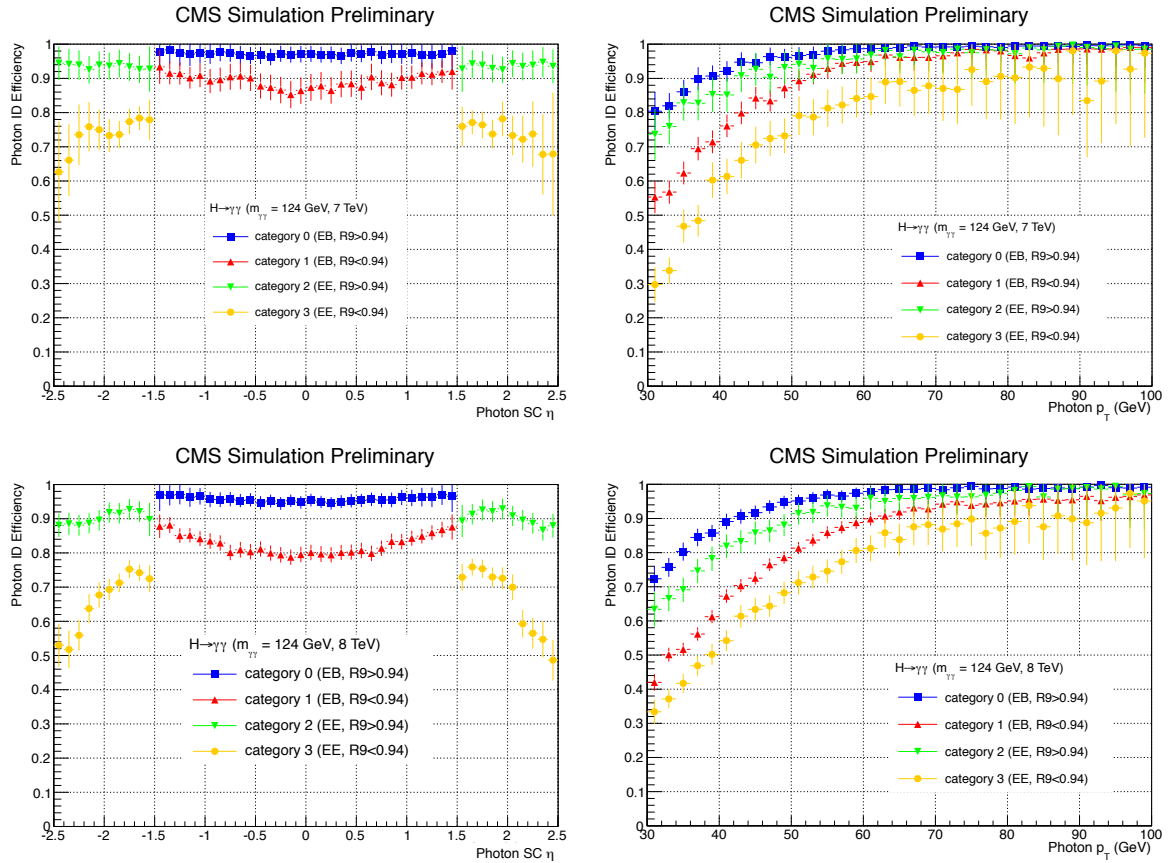
- PF isolation sum, chosen vertex - Sum of particle flow photon, charged and neutral hadron isolation sums defined in Sec. 3.5 where all PF candidates originate from the primary vertex selected in Sec. 4.4
- PF isolation sum, worst vertex - As above but where all PF candidates originate instead from the vertex which gives the largest charged hadron PF isolation sum. This adds protection for cases when the primary vertex is incorrectly assigned.
- Charged PF isolation sum - Described above in Sec. 3.5.
- $\sigma_{i\eta i\eta}$  - Described above in Sec. 4.5.
- $H/E$  - Described above in Sec. 4.5.
- $R_9$  - Described above in Sec. 3.2.2.

### 5.1.2 Photon ID MVA

For the MFM and SMVA analyses a BDT is trained to discriminate between prompt photons and jets. The desire is to factorise the photon selection, which is required to distinguish prompt photons from neutral mesons faking photons, and the event selection, which is required to consider kinematics, resolution, etc. to distinguish  $H \rightarrow \gamma\gamma$  from the  $pp \rightarrow \gamma\gamma$ ,  $\gamma$ +jet, jet + jet background. The input variables used are designed specifically to distinguish between photons and fakes. They should not have any properties which make the identification Higgs specific. Consequently, the training samples used are  $\gamma$ +jet

	Barrel		Endcap	
	$R_9 \geq 0.94$	$R_9 < 0.94$	$R_9 \geq 0.94$	$R_9 < 0.94$
PF isolation sum, chosen vertex (GeV)	<6	<4.7	<5.6	<3.6
PF isolation sum, worst vertex (GeV)	<10	<6.5	<5.6	<4.4
Charged PF isolation sum (GeV)	<3.8	<2.5	<3.1	<2.2
$\sigma_{i\eta i\eta}$	<0.0108	<0.0102	<0.028	<0.028
H/E	<0.124	<0.092	<0.142	<0.063
$R_9$	$\geq 0.94$	$> 0.298$	$\geq 0.94$	$> 0.24$

**Table 5.1:** Photon ID selection cut values. The cuts are applied to both the leading and subleading photons.



**Figure 5.1:** Cut based photon ID efficiency as measured in  $Z \rightarrow e^+e^-$  MC simulation using the tag and probe method[57]. Shown for the 7 TeV dataset (top row) and 8 TeV dataset (bottom row) as a function of the photon supercluster position in  $\eta$  (left) and as a function of the photon  $p_T$  (right). Photons in the barrel which are unconverted (converted) are plotted as the blue (red) points and photons in the endcap which are unconverted (converted) are plotted as the green (yellow) points.



samples where the identification BDT signal is the prompt  $\gamma$  and the background is the fake jet. The  $p_T$  and supercluster  $\eta$  distributions are reweighted such that they match between signal and background. This negates the BDT exploiting any photon kinematics which could correlate to the Higgs mass. The photon ID BDT is trained separately for the barrel and endcap as these regions of phase space are so different. Separate BDTs were also trained for 7 and 8 TeV. The result is four separate trainings in total.

The input variables aim to exploit differences in the shower shape and isolation between prompt and non-prompt photons and the correlation between these variables and the supercluster position and energy. They are:

### Shower shape variables

- $\sigma_{i\eta i\eta}$  - Explained above in Sec. 4.5.
- $\sigma_{i\eta i\phi}$  - The equivalent diagonal spread (in  $\eta, \phi$ ), representing the  $(\eta, \phi)$  correlation of the shower.
- $E_{2 \times 2} / E_{5 \times 5}$  - Ratio of the energy in the most energetic  $2 \times 2$  cluster which contains the seed to the energy in the  $5 \times 5$  cluster.
- $R_9$  - Explained above in Sec. 3.2.2.
- $\sigma_\eta$  - The energy-weighted standard deviation of single crystal  $\eta$  within the super-cluster.
- $\sigma_\phi$  - The energy-weighted standard deviation of single crystal  $\phi$  within the super-cluster.
- $\sigma_{xy}$  (for endcap only) - The standard deviation of the shower spread in the  $x, y$  planes of the preshower, representing the  $x$ - $y$  correlation of the shower.

### Isolation variables

- PF Photon ISO - Particle flow photon isolation sum.
- PF Charged Hadron ISO (selected vertex) - Particle flow charged hadron isolation sum for candidates originating from selected vertex.
- PF Charged Hadron ISO (worst vertex) - Particle flow charged hadron isolation sum for candidates originating from the vertex with the largest isolation sum.

### Correlation variables

- $\rho$  - The median energy density in the event.
- $\eta$  - The  $\eta$  position of the photon supercluster.
- $E_{raw}$  - The raw energy of the photon supercluster.

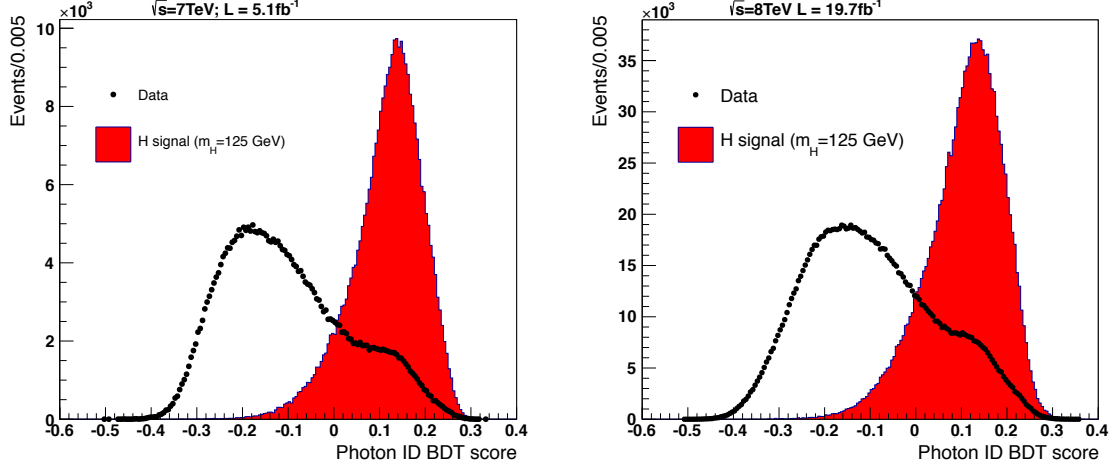
The testing sample used to verify the output of the photon identification BDT is a MC  $H \rightarrow \gamma\gamma$  sample ( $m_H = 124$  GeV). The photon identification BDT output provides a measure of an individual photon’s “quality” and is used as an input to the event-level BDT (described in the next section). Even so, a considerable amount of background is cut out by defining a loose cut (the BDT output must be  $> -0.2$ ) on the photon ID BDT output value which is more than 99% signal efficient. This is demonstrated in Fig. 5.2 which shows the photon ID BDT output for a  $H \rightarrow \gamma\gamma$  signal MC sample and for all the data events which pass the preselection defined in Sec. 4.5. It is clear that placing a cut at  $> -0.2$  removes a considerable amount of the background whilst maintaining very high signal efficiency.

The photon ID BDT response for each photon is used as a direct input to the event-level MVA. Given that imperfect modeling of the detector response can result in a small change in the photon ID response, which has a direct impact on the event-level MVA response, which is used to classify events, a systematic error on the photon quality is applied and propagated through to the event-level MVA. Validation of the photon ID BDT response in the  $Z \rightarrow e^+e^-$  decay is shown in Fig. 5.3 with the size of the systematic error applied to account for any data/MC discrepancies. The uniform response of the identification as a function of the number of primary vertices is demonstrated by the similarity of the two plots in this figure (Fig. 5.3) which are for events in which the number of primary vertices is  $\leq 15$  (on the left) and for those in which the number of primary vertices is  $> 15$  (right).

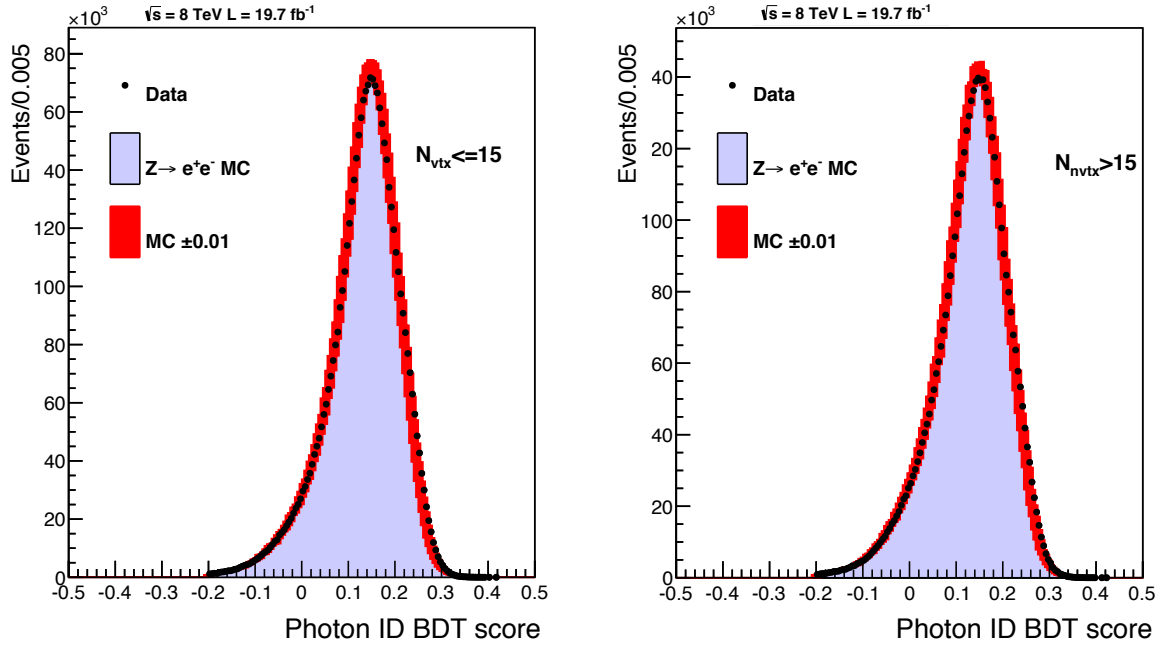
### 5.1.3 Diphoton event-level MVA

Whilst the CiC analysis selects events based on photon identification, the MVA analysis approach is to first select photons using the photon identification BDT described in the section above and then pass all the relevant event information through an event-level BDT. The event-level classifier, referred to as the diphoton BDT, is constructed to give a high score to events which fulfill the following criteria:

1. The event kinematics should be compatible with a Higgs decay.



**Figure 5.2:** The output distribution of the photon identification BDT for 7 TeV (left) and 8 TeV (right) datasets. The black points show the output for data events which pass the preselection and the red histogram shows the output for a simulated MC sample of  $H \rightarrow \gamma\gamma$  decays. A cut of  $> -0.2$  is made on all photons.



**Figure 5.3:** The output distribution of the photon identification BDT for the 8 TeV training as validated by the  $Z \rightarrow e^+e^-$  decay for events which have  $\leq 15$  reconstructed primary vertices (left) and those which have  $> 15$  primary vertices (right). The data is shown as the black points with the MC simulation as the blue histogram. The systematic uncertainty on the output as applied to the MC sample is shown as the red band.

2. The event has good mass resolution.
3. The event contains two “high quality” photons (i.e. they have a high score from the photon ID BDT).

It is highly important that the BDT is completely independent of Higgs mass and consequently that the input variables have no, or at the least very little, dependence on the Higgs mass. This is essential to have a fair training. If the BDT included the Higgs mass, or a variable highly correlated with it, it would preferentially select events with this mass therefore biasing the selection towards events which have a mass near the mass of the signal used to train with. The input variables used are:

### Event kinematics

- $p_T^{1(2)}/m_{\gamma\gamma}$  - The mass relative transverse momenta of each photon.
- $\eta^{1(2)}$  - The pseudorapidity of each photon.
- $\cos(\phi_1 - \phi_2)$  - The cosine of the angle between the two photons in the transverse plane. This variable reflects the  $p_T$  of the diphoton system (in other words the reconstructed Higgs candidate) without introducing a mass dependence.

### Mass resolution

- $\sigma_m^{right}/m_{\gamma\gamma}$  - The mass resolution of the event assuming the correct primary vertex has been selected. In this case the angular resolution is negligible so the variable is calculated using just the two photon energy resolution values as

$$\frac{\sigma_m^{right}}{m_{\gamma\gamma}} = \frac{1}{2} \left( \frac{\sigma_{E_1}}{E_1} \oplus \frac{\sigma_{E_2}}{E_2} \right). \quad (5.1)$$

- $\sigma_m^{wrong}/m_{\gamma\gamma}$  - The mass resolution of the event assuming the wrong vertex is selected. The vertex position in  $z$  is distributed as a Gaussian with a width equivalent to  $\sqrt{2}\sigma_z^{beamspot}$  and so the angular resolution  $\sigma_m^{vtx}$  can be analytically calculated given the ECAL impact positions of the two photons. Consequently the wrong vertex variable is calculated as

$$\frac{\sigma_m^{wrong}}{m_{\gamma\gamma}} = \frac{\sigma_m^{right}}{m_{\gamma\gamma}} \oplus \frac{\sigma_m^{vtx}}{m_{\gamma\gamma}}. \quad (5.2)$$

- $p_{vtx}$  - The probability that the selected primary vertex is correct. In order to tie together the mass resolution information given the right vertex hypothesis and the wrong vertex hypothesis, the probability that the vertex is correct is used in addition.
- It is also important to specify in the training that the signal to background ratio is inversely proportional to the mass resolution. Accordingly the signal events in the training are weighted by a factor,

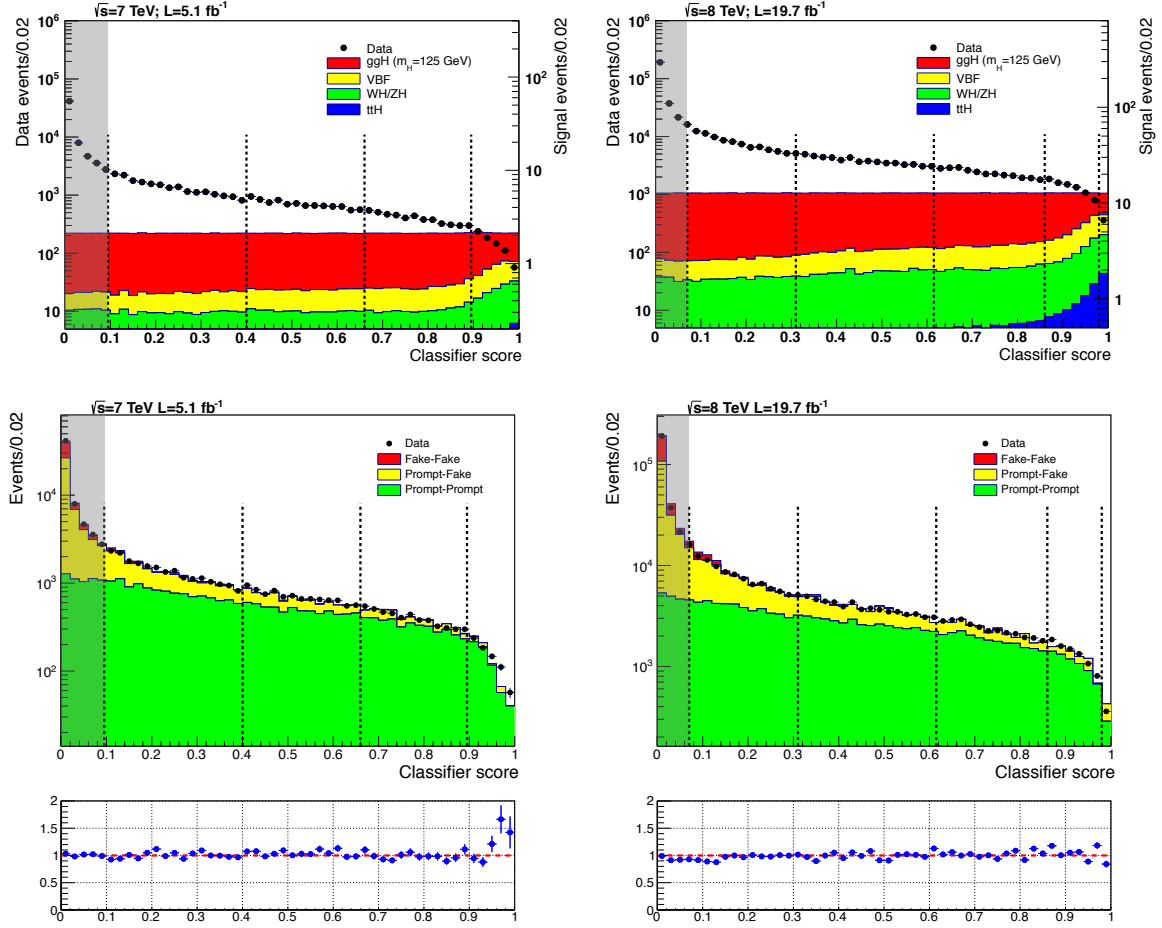
$$w = \frac{p_{vtx}}{\sigma_m^{right}/m_{\gamma\gamma}} + \frac{1 - p_{vtx}}{\sigma_m^{wrong}/m_{\gamma\gamma}}. \quad (5.3)$$

### Photon quality

- $\text{phoID}^{1(2)}$  - The photon ID BDT output value of each photon.

The training is performed separately for 7 and 8 TeV and the samples used for the signal are all of the SM  $H \rightarrow \gamma\gamma$  MC samples ( $ggH$ ,  $VBF$ ,  $VH$ ,  $t\bar{t}H$ ) appropriately weighted by cross section and the samples used for the background are the cross section-weighted mixture of SM backgrounds which include contributions from  $pp \rightarrow \gamma\gamma$  (prompt-prompt),  $pp \rightarrow \gamma + \text{jet}$  (prompt-fake) and  $pp \rightarrow \text{jet} + \text{jet}$  (fake-fake) as described in Sec. 4.2.1. The training is performed on only half of each event sample (selected by even event number) so that the BDT response can be tested on the other half (selected by odd event number).

A cut is placed on the BDT response in order to remove almost all of the events which contain two fake photons and a large fraction of those which contain one fake. The remaining events which pass this cut are those used in the results and get categorised in coarse bins based on the BDT response. The strategy for optimising this cut value and the category boundaries is explained in Sec. 5.2. The BDT response in data, background and signal is shown in Fig. 5.4. The BDT output in this figure has been transformed so that its response is flat for the total signal. This helps to see the differences between the signal (flat) and background (not flat) and also the differences between the gluon fusion signal and the other production modes. By examining the BDT response in signal (top row of Fig. 5.4) one can see that the associated production modes ( $VBF$  in yellow,  $VH$  in green and  $t\bar{t}H$  in blue) peak closer to 1 relative to the gluon fusion signal (red). This is predominantly driven by the fact that Higgs bosons produced by associated modes are typically at higher  $p_T$  than for gluon fusion and this kinematic feature helps the BDT to discriminate these events from the background. It can be seen in the bottom row of Fig. 5.4 that the cut on the event-level BDT removes a considerable amount of the



**Figure 5.4:** The diphoton BDT response for the 7 TeV training (left column) and 8 TeV training (right column) transformed so that the output is flat in the signal. The top row shows the output in data (black points), which contains mostly background events, and signal MC events (filled histograms) split by production mode;  $ggH$  (red),  $VBF$  (yellow),  $VH$  (green),  $t\bar{t}H$  (blue). The bottom row shows the output in data (black points) and background MC events (filled histograms) split by type; prompt-prompt (green), prompt-fake (yellow) and fake-fake (red). The data/background residual is shown as the blue points underneath these histograms. The vertical dashed lines show the analysis category definitions where those on the right (nearer a classifier score of 1) have the highest S/B ratio. All events which fall below the left most boundary, shown by the shaded area, are cut out of the analysis.

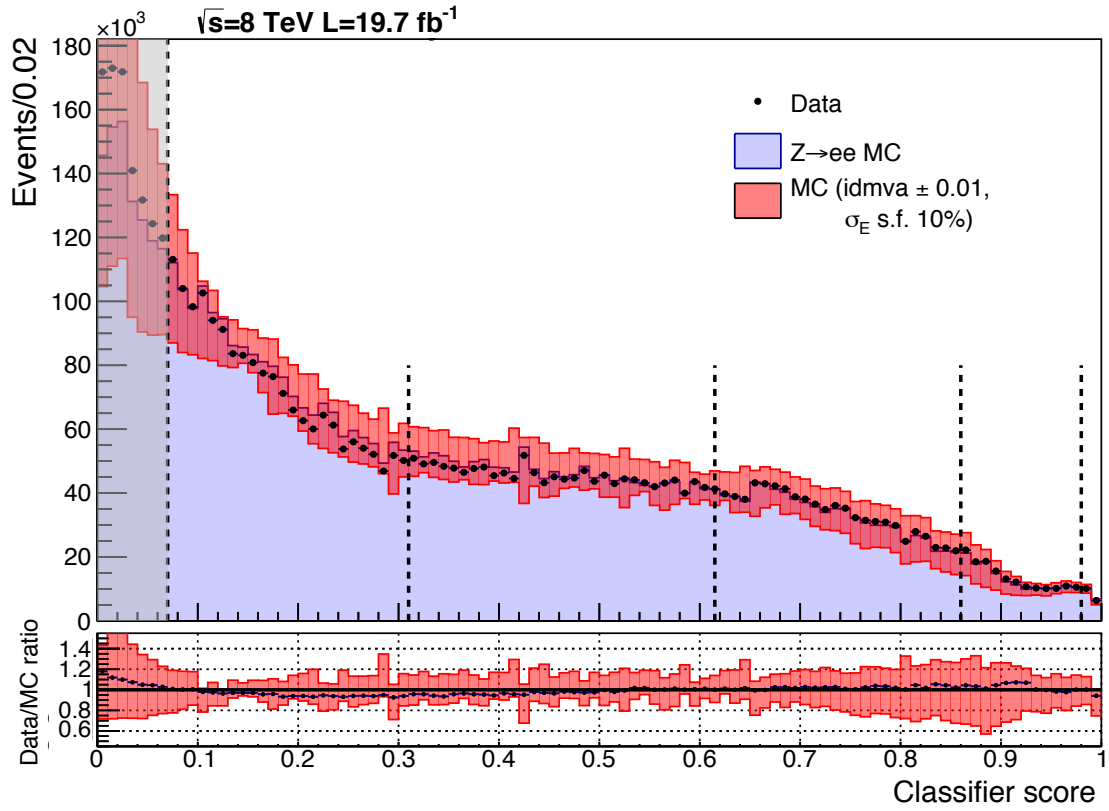
fake-fake and prompt-fake contribution to the background with the remainder consisting of about 70% prompt-prompt and 30% prompt-fake.

Any uncertainties which affect the shape of the output distribution of this BDT result in event migrations between the final analysis categories; if the BDT response is mismodelled then events can move across the boundaries represented as the vertical

dashed lines in Fig. 5.4. The signal model in this analysis is obtained from MC simulation and so these uncertainties can have an effect on the signal model shape in each of the categories. Whilst these migrations may also be true of the background the effect is not important because the background is extracted from data. The input variables whose uncertainties have the largest effect on the BDT response in signal are the photon ID quality and the photon energy resolution estimate. This is because these variables have both i) relatively large uncertainties because of imperfect detector response modelling in the simulation and ii) are highly discriminative and hence can have a relatively large impact on the BDT response. As the BDT response varies monotonically with these variables, the systematic uncertainty on them is propagated through the analysis as an event migration. Systematic uncertainties are described in more detail in Sec. 6.3. The size of this effect is shown in the BDT response validation plot using the  $Z \rightarrow e^+e^-$  decay in Fig. 5.5.

## 5.2 Event categorisation

In order to exploit different regions of phase space with dissimilar signal to background ratios the events are split into categories. Furthermore, additional categories can be designed to enrich the selection with events characteristic of particular Higgs production modes. The  $VBF$  production mode is typically accompanied by a pair of jets with a large pseudorapidity separation. The  $VH$  production, which includes contributions from  $WH$  and  $ZH$ , may be accompanied by a charged lepton, missing transverse energy or jets originating from the decay of the associated  $W$  or  $Z$  boson. Similarly,  $t\bar{t}H$  production may be accompanied by  $b$  quarks and/or charged leptons. The predominant production mode, which accounts for about 88% of the signal, is  $ggH$  and is inclusive. The amount of signal produced from exclusive modes is approximately 8% for  $VBF$ , 4% for  $VH$  and <1% for  $t\bar{t}H$ . By including a series of event tags, and separating events accordingly, all four production modes of the Higgs at the LHC are harnessed in this analysis. This not only helps to increase the overall sensitivity to a SM Higgs boson (given the very low background rates expected for the exclusive modes) but also significantly helps to reduce the error on measurements of the relative couplings of the observed boson to fermions and bosons, as the signal from the relative production modes gets split into distinct categories. This “exclusive mode tagging” is done identically for both the nominal MFM analysis and also the cross-check SMVA, although the inclusive mode categorisation is done differently. The CiC spin analysis uses no exclusive mode tagging as the spin-2



**Figure 5.5:** The diphoton BDT response for the 8 TeV training in the  $Z \rightarrow e^+e^-$  decay. The data is shown as the black points and the MC events as the blue histogram. The systematic applied to account for variation in the BDT response from mismodelling in the photon quality response and the photon energy resolution estimate are shown as the red band. The vertical dashed lines show the boundaries of the analysis categories.



models considered (minimal coupling graviton) have only inclusive production modes. There is an alternate categorisation scheme used for the spin analysis which exploits both the event mass resolution and the differential decay angle.

### 5.2.1 Exclusive mode tagging

All events which make it to this stage will have passed the photon preselection (Sec. 4.5), as well as the basic requirement of two high  $p_T$  photons with  $100 < m_{\gamma\gamma} < 180$  GeV (Sec. 5.1) and the MFM analysis selection (Sec. 5.1.3), which includes both the photon ID MVA cut and the diphoton event-level BDT cut. These make up the final event sample. They now pass through the “tagging” procedure described below in which they are organised into a set of non-overlapping event classes. The tagging is done in a specific order to ensure there is no overlap between classes and the order is chosen such that preference is given to categories with a higher expected signal to background ratio. The aim of the exclusive mode tagging is to separate events into their most probable production mode (one of  $VBF$ ,  $VH$  or  $t\bar{t}H$ ) where the events which are left “Untagged” are mostly produced by  $ggH$ . The tagging order is shown alongside a summary of the relevant cut values in Table 5.4 at the end of the chapter. If an event does not meet the requirements of a particular tag it is passed onto the next tag and if it fails all tag requirements it is placed in one of the inclusive categories, whose structure is described below (Secs. 5.2.2-5.2.3), meaning that no event cannot be included in the analysis.

#### Dijet-tagged categories for VBF

The following variables are used to exploit the specific topology of the jet pairs associated to  $VBF$  Higgs production:

- $p_T^\gamma/m_{\gamma\gamma}$  - The transverse momenta of the leading and subleading photons as a ratio of the diphoton invariant mass.
- $p_T^j/m_{\gamma\gamma}$  - The transverse momenta of the leading and subleading jets as a ratio of the diphoton invariant mass.
- $m_{jj}$  - The dijet invariant mass.
- $|\Delta\eta_{j_1j_2}|$  - The absolute pseudorapidity difference between the two jets.
- $Z = \eta(\gamma_1) + \eta(\gamma_2) - [\eta(j_1) + \eta(j_2)]/2$  - The so-called *Zeppenfeld* variable[58].

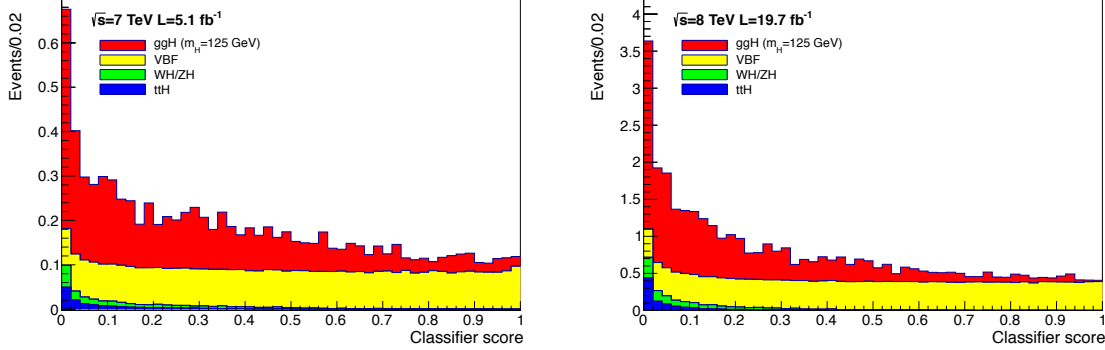
- $\Delta\phi_{j_1j_2}$  - The angular difference between the two jets in the transverse plane.

Additionally the leading photon  $p_T$  requirement is raised to  $p_T^{\gamma_1}/m_{\gamma\gamma} > 1/2$ . The energy measurement of jets in the event are calibrated to correct for detector effects[59] and additional energy in the jets from pileup is removed using the FASTJET jet areas technique described in[60, 61, 62]. Jets are required to be within the pseudorapidity range  $|\eta| < 4.7$ .

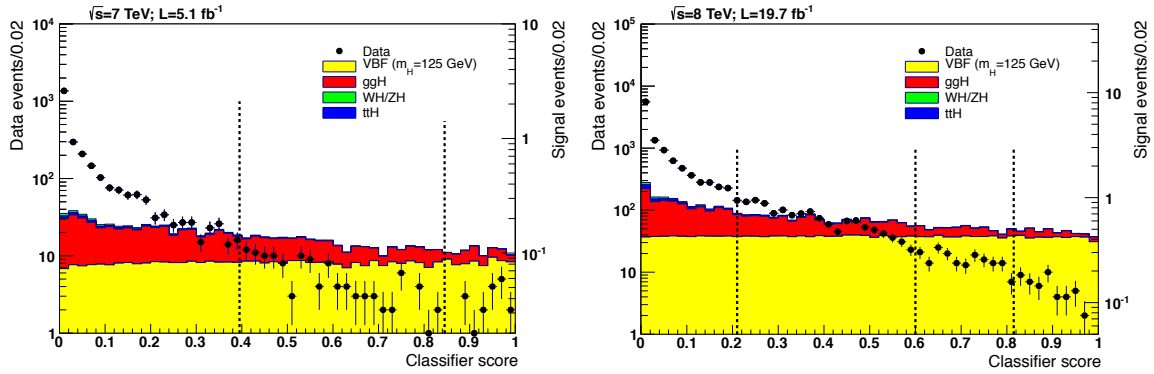
The dijet tagging is done with use of two additional MVAs. The first is designed to exploit the  $VBF$  kinematic properties and the second is used to combine this information with the diphoton BDT. Candidates are required to pass a  $VBF$  preselection of two jets with  $p_T^{j_1} > 30$  GeV and  $p_T^{j_2} > 20$  GeV and invariant mass,  $m_{jj} > 75$  GeV. The signal sample used for training is the SM MC simulation with just  $VBF$  production, whilst the SM gluon fusion MC simulation is included as background along with the usual prompt-prompt, prompt-fake and fake-fake contributions. This helps produce an output in which a high score gives a very pure  $VBF$  sample.

The combined dijet-diphoton BDT has inputs of the kinematic dijet BDT output, the diphoton BDT output and  $p_T^{\gamma\gamma}/m_{\gamma\gamma}$  in order to discriminate  $VBF$  from both the other signal types and the background, utilising all the information available in the event. The transverse momenta of the diphoton system as a ratio of the diphoton invariant mass,  $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ , is included both because of its discriminating power and the differences between signal and background of its correlation with both the dijet BDT output and the diphoton BDT output.

One finds that the background rejection for  $VBF$  is significantly improved by the use of the combined dijet-diphoton BDT whilst the  $VBF$  purity (i.e. separation from  $ggH$ ) is not as good when collapsing the kinematic dijet BDT and the combined BDT into one step. Consequently the trainings are performed separately and the  $VBF$  categories are defined by picking out regions which have a high score in the combined dijet-diphoton BDT response. Each successive BDT training uses statistically independent MC samples to avoid selection bias from fluctuations in the simulation. The optimisation procedure for deciding the category boundaries is analogous to the one used for the inclusive categories in the MFM analysis where the target is to minimise the expected uncertainty of the signal strength from the  $VBF$  process alone when moving the boundaries around. The optimisation procedure includes the statistical disadvantage of having too many categories for a given dataset size. This is explained further in Section 5.2.2. At 8 TeV there are three  $VBF$  categories and at 7 TeV there are two, because of the considerably lower



**Figure 5.6:** The distributions of the kinematic dijet BDT response at 7 TeV (left) and 8 TeV (right) for the signal split by production mode;  $ggH$  (red),  $VBF$  (yellow),  $VH$  (green) and  $t\bar{t}H$  (blue). The output is transformed so that it is flat in the  $VBF$  signal.



**Figure 5.7:** The distributions of the combined dijet-diphoton BDT response at 7 TeV (left) and 8 TeV (right) for the data (black points), which contains mostly background, and signal (filled histograms) split by production mode;  $ggH$  (red),  $VBF$  (yellow),  $VH$  (green) and  $t\bar{t}H$  (blue). The output is transformed so that it is flat in the  $VBF$  signal. The vertical dashed lines show the  $VBF$  category definitions where those on the right (nearer a classifier score of 1) are purest  $VBF$  signal. All events which fall below the left most boundary are not included in  $VBF$  categories but fall back to being categorised amongst the inclusive categories.

statistics in the 7 TeV dataset. The output distributions at 7 and 8 TeV for the signal are shown for the kinematic dijet BDT in Fig. 5.6 and for the combined diphoton-dijet BDT, along with the  $VBF$  category boundaries, in Fig. 5.7. The BDT response shown in these plots is transformed so that it is flat in the  $VBF$  signal (yellow histogram). One can see that the other signal types peak lower in the transformed score and the data, which contains mostly background, even more so.

### Lepton-, jet- and $\cancel{E}_T$ -tagged categories for $VH$

The selection for the four categories designed to tag  $VH$  production are optimised by minimising the expected uncertainty on the signal strength of this process alone. Two of the classes require at least one charged muon or electron and are split into a tight selection category and a looser selection category, the third is for events consistent with large  $\cancel{E}_T$  and the fourth for events consistent with two or more jets. The leading photon cut is raised to  $p_T^{\gamma_1}/m_{\gamma\gamma} > 3/8$  for all  $VH$ -tagged categories. The category requirements are as follows:

- **$VH$  Tight  $l$ :** The tightly selected lepton category is characterised by the signature of a leptonically decaying  $W$  or  $Z$  boson and as such requires the presence of  $\cancel{E}_T > 35$  GeV or another lepton of the same flavour and opposite charge as the first. In the first case (of a lepton +  $\cancel{E}_T$ ) the lepton is required to have  $p_T > 20$  GeV, in the latter case (of two leptons) the requirement is  $p_T > 10$  GeV for both leptons whilst the invariant mass of the dilepton pair must be  $70 < m_{ll} < 110$  GeV. The diphoton BDT output is required to be  $> 0.1$  ( $> -0.6$ ) for the 7 (8) TeV datasets.
- **$VH$  Loose  $l$ :** For the loosely selected lepton category the lepton  $p_T$  must satisfy  $p_T > 20$  GeV. The selection requirements are designed to reduce the background from leptonic  $Z$  bosons (not associated with a Higgs) that contain initial or final state radiation faking the diphoton signal. Consequently, leptons are required to be separated by at least  $\Delta R > 1.0$  from the closest photon and the invariant mass of any lepton-photon pairs must be more than 10 GeV away from the  $Z$  boson mass. In addition a conversion veto is applied to the electrons to reduce the rate of misidentified photons. The diphoton BDT output is required to be  $> 0.1$  ( $> -0.6$ ) for the 7 (8) TeV datasets.
- **$VH$   $\cancel{E}_T$  tag:** Accurate measurement and simulation of the  $\cancel{E}_T$  vector has been studied in detail at CMS and a set of standard corrections (for both data and simulation) are applied[63]. The corrected  $\cancel{E}_T$  is required to pass  $\cancel{E}_T > 70$  GeV whilst the angular separation in the transverse plane between the momentum of the diphoton system and the  $\cancel{E}_T$  direction must pass  $\Delta\phi(\gamma\gamma, \cancel{E}_T) > 2.1$ , and similarly the angle between the momentum of the diphoton system and the leading jet must pass  $\Delta\phi(\gamma\gamma, \text{jet}) < 2.7$ . The diphoton BDT output is required to be  $> 0.8$  ( $> 0.0$ ) for the 7 (8) TeV datasets.

- **$VH$  dijet tag:** The event must contain at least one jet pair in which both jets have  $p_T > 40$  GeV and  $|\eta| < 2.4$  and have an invariant mass within the range  $60 < m_{jj} < 120$  GeV. The diphoton transverse momentum must satisfy  $p_T^{\gamma\gamma}/m_{\gamma\gamma} > 13/12$ . Additionally the angular correlation between the diphoton system and the dijet system from  $VH$ -associated Higgs decays can be exploited. The angle,  $\theta^*$ , between the diphoton direction in the diphoton-dijet rest frame and the lab frame is flat for events from  $VH$  decays whereas for the background and gluon fusion-produced Higgs decays the distribution peaks at  $|\cos(\theta^*)| = 1$ . Consequently, there is a requirement that  $|\cos(\theta^*)| < 0.5$ . The diphoton BDT output is required to be  $> 0.6$  ( $> 0.2$ ) for the 7 (8) TeV datasets.

### Lepton- and jet-tagged categories for $t\bar{t}H$

There are two categories for tagging production from  $t\bar{t}H$  decays, one of which is lepton-based and one of which is jet-based. The total fraction of signal expected from  $t\bar{t}H$  is  $< 1\%$  so only a handful of events are expected. Consequently for the 7 TeV dataset the two categories are merged into one class. As for the  $VH$ -tagged categories, the cuts are optimised to minimise the expected uncertainty of the signal strength measurement of the  $t\bar{t}H$  process alone.

For both classes the leading photon  $p_T$  cut is raised to  $p_T^{\gamma_1}/m_{\gamma\gamma} > 1/2$ , all jets are required to have  $p_T > 25$  GeV and there must be at least one b-tagged jet present. The specific requirements of each category are as follows:

- **$t\bar{t}H$  multijet tag:** The requirement is at least four additional jets in the event and no lepton. The diphoton BDT output is required to be  $> 0.6$  ( $> -0.2$ ) for the 7 (8) TeV datasets.
- **$t\bar{t}H$  lepton tag:** At least one more jet in the event and one muon or electron which has  $p_T > 20$  GeV. The diphoton BDT output is required to be  $> 0.6$  ( $> -0.6$ ) for the 7 (8) TeV datasets.

### 5.2.2 Inclusive mode categorisation and VBF dijet categorisation in the mass-factorised MVA analysis

The combined dijet-diphoton BDT output value is used to define a set of  $VBF$  categories. The goal is to find the configuration of category boundaries which minimise the expected

uncertainty on the signal strength for  $VBF$  production alone, allowing the number of categories and where the category boundaries lie to be completely free floating, with the additional requirement that the efficiency  $\times$  acceptance of the categories matches between 7 and 8 TeV. This results in a tight  $VBF$  category for events with a very high combined dijet-diphoton BDT score, a somewhat looser category and then one or more very loose categories. It is found that dropping the loosest category has a negligible impact ( $<1\%$ ) on the expected uncertainty and as such the upper boundary for the loosest category is turned into a lower cut. All events which pass the  $VBF$  preselection (described in Sec. 5.2.1) and fail the lower cut are then classified somewhere in the inclusive categories defined below.

Once the  $VBF$  category boundaries have been found the same procedure is deployed using the diphoton BDT output value. This time the target is to minimise the expected uncertainty on the total signal strength, allowing the number of categories and the category boundary values to be completely free floating. One finds a rather similar structure and that dropping events in the very loosest category has a negligible impact on the performance and consequently this dictates the lower cut value for the diphoton BDT. The expected uncertainty minima, when floating both the number of category boundaries and the boundary locations, are fairly shallow so the exact position of the boundaries has a very small impact on the performance of the analysis. For the 7 (8) TeV datasets there are 4 (5) inclusive categories with a lower diphoton BDT cut of 0.19 ( $-0.78$ ) and 2 (3)  $VBF$  categories<sup>1</sup>.

### 5.2.3 Inclusive mode categorisation in the sideband MVA analysis

In the SMVA analysis all the exclusive categories are identical to the MFM analysis (including the  $VBF$  categories). However the inclusive categorisation is done slightly differently. In the sideband analysis the signal region is defined in a  $\pm 2\%$  window around the hypothesis Higgs mass and thus contains  $\sim 75\%$  of the signal. The analysis is performed as a cut and count in the signal window over several bins. There is one bin for each exclusive category and then several more for the inclusive. The binning scheme for the inclusive events is defined as follows,

---

<sup>1</sup>There are fewer categories in the 7 TeV dataset because there are less statistics.

- Make two-dimensional distributions of the diphoton BDT score and the distance of the invariant mass from the hypothesised Higgs mass,  $\Delta m/m_H$ , in the  $\pm 2\%$  window for signal and background, as shown in Figure 5.8, where,

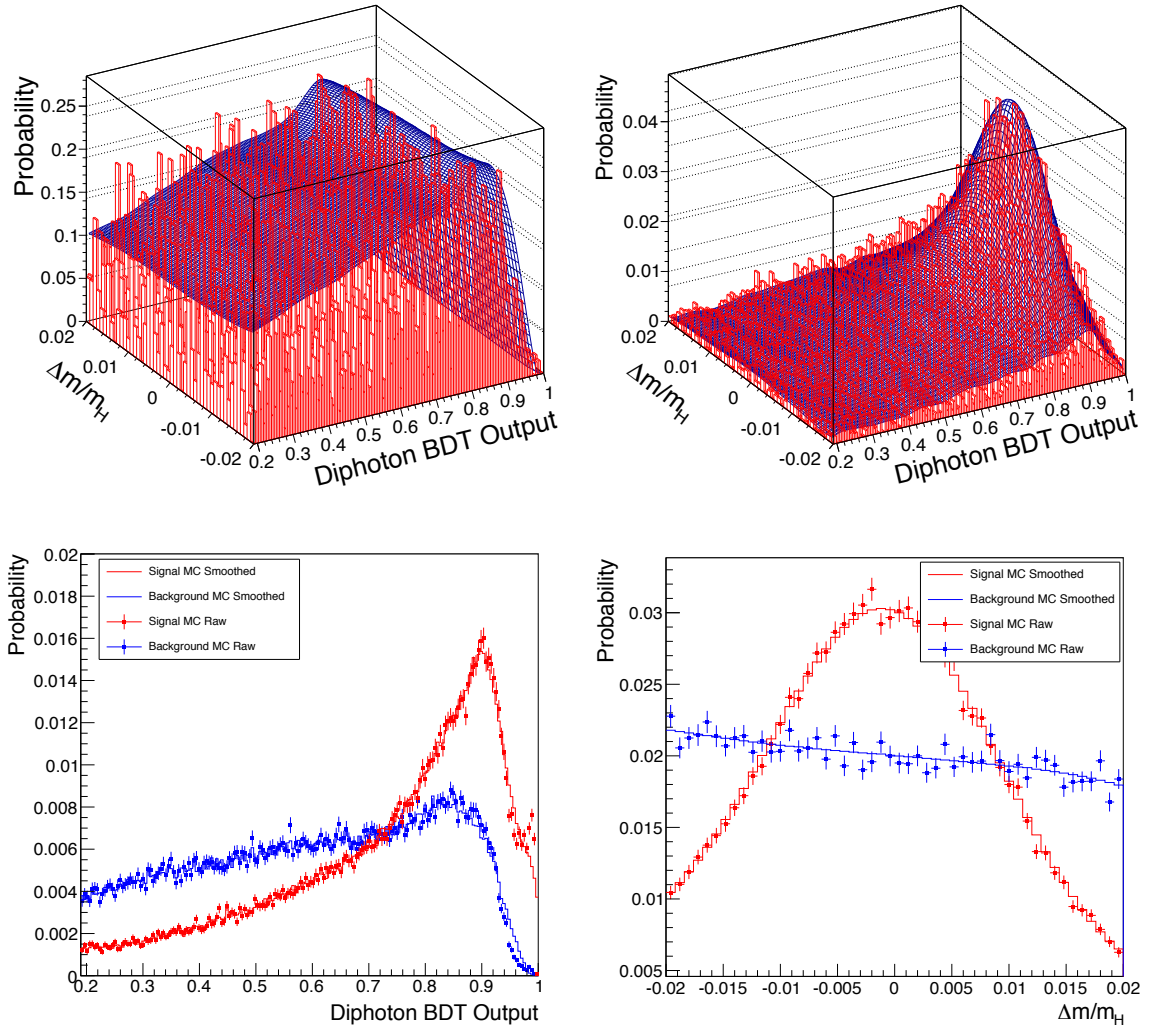
$$\frac{\Delta m}{m_H} = \frac{m_{\gamma\gamma} - m_H}{m_H}. \quad (5.4)$$

- Use a local regression smoothing technique[64] to smooth statistical fluctuations within the sample (demonstrated in Fig. 5.8).
- Select bins by isolating regions of this 2D phase space which have similar  $S/B$  ratios and optimise the boundaries to give the maximum expected signal significance.

Clearly the most sensitive bins will be the ones which have a high diphoton BDT score and have a low value of  $|\Delta m/m_H|$  (i.e. are near the signal peak). The category boundaries in this 2D plane are shown as different shades in Figure 5.9. In total there are 8 (10) inclusive bins for the 7 (8) TeV samples in the SMVA. There are some small residual differences between the raw MC distributions and the smoothed distributions, shown in Fig. 5.8, especially near the high side boundary where smoothing algorithms can have difficulties. However, it has been found that artificially resolving such discrepancies results in a small change to the position of the category boundaries and furthermore that small changes in the boundary definitions have a very small impact on the sensitivity of the analysis. It should be noted that changing the location of the category boundaries cannot bias the result but only provide a non-optimal result. Consequently such discrepancies are considered unimportant as they do not influence the central value of the result and are only estimated to decrease the sensitivity by  $< 2\%$ .

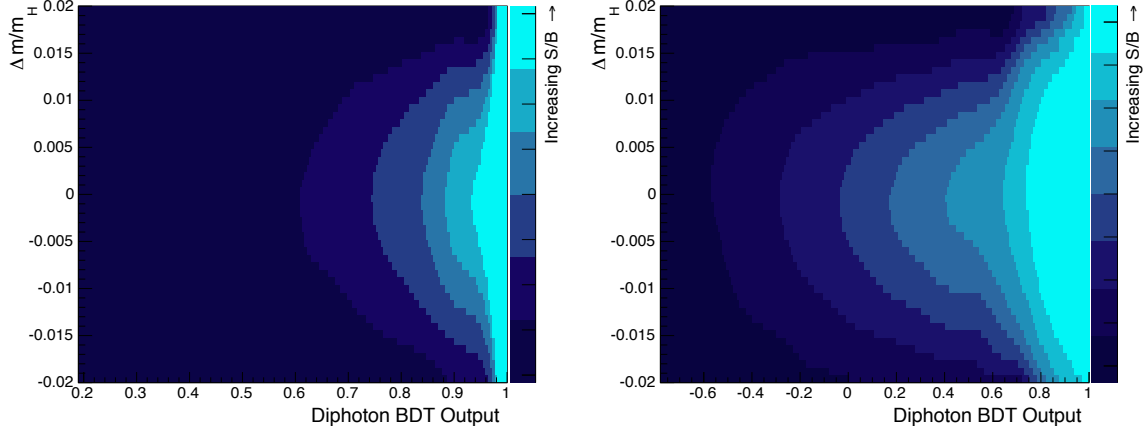
### 5.2.4 Inclusive mode categorisation in the spin analysis

The Landau-Yang theorem forbids the direct decay of a spin-1 particle into a pair of photons[65, 66]. Consequently the spin analysis compares the expectation of the spin-0 SM Higgs,  $J^P = 0^+$ , and the spin-2 *graviton-like* model with minimal couplings,  $2_m^+$ [67]. The  $2_m^+$  graviton resonance is produced in one of two ways, gluon-fusion ( $gg$ ) or quark-antiquark annihilation ( $q\bar{q}$ ). As the  $2_m^+$  is just one of many spin-2 models it is desirable to make the analysis as model-independent as possible. As a means of discriminating between the two hypotheses, use is made of the scattering angle in the Collins-Sopfer frame,  $\cos(\theta_{CS}^*)$ [68], which is defined as the angle, in the diphoton rest frame, between



**Figure 5.8:** Two dimensional distributions of the diphoton BDT output and  $\Delta m/m_H$  are shown on the top row for the background (left) and signal (right) for the 7 TeV sample. The bottom rows shows the projections for signal (red) and background (blue) in the two variables.





**Figure 5.9:** The inclusive category bin definitions for the SMVA analysis. Shown for the 7 TeV dataset on the left and the 8 TeV dataset on the right.

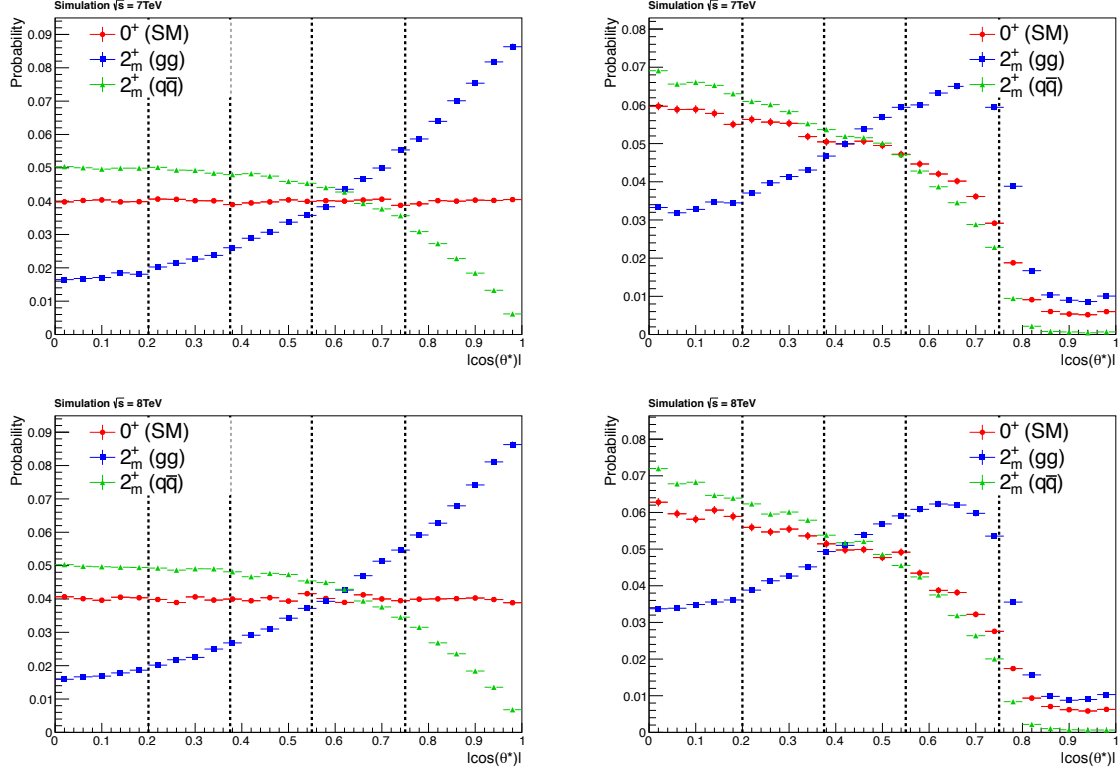
the collinear diphotons and the line which bisects one incoming beam with the negative of the other beam,

$$\cos(\theta_{\text{CS}}^*) = 2 \times \frac{E_2 p_{z1} - E_1 p_{z2}}{m_{\gamma\gamma} \sqrt{m_{\gamma\gamma}^2 + p_{T\gamma\gamma}^2}}, \quad (5.5)$$

where  $E_1$  and  $E_2$  are the energies of the leading and trailing photon,  $p_{z1}$  and  $p_{z2}$  are the  $z$ -component momenta of the leading and trailing photon and  $m_{\gamma\gamma}$  and  $p_{T\gamma\gamma}$  are the invariant mass and transverse momenta of the diphoton system.

In the diphoton rest frame the photons from the decay of a spin-0 boson are isotropic. Hence prior to acceptance cuts, the distribution of  $\cos(\theta_{\text{CS}}^*)$  under the  $0^+$  hypothesis is uniformly flat. In general this is not the case for spin-2 decays.

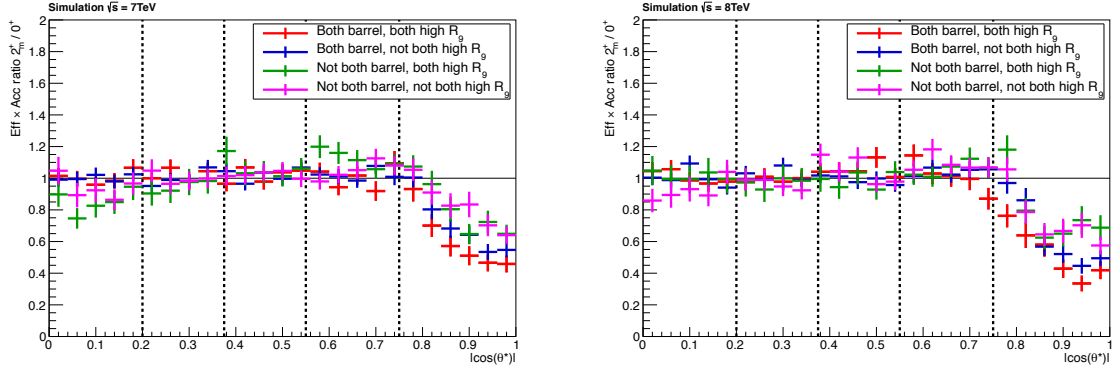
In order to reduce any model dependence in the analysis the cut-based photon selection, described in Sec. 5.1.1, is used to pick events by applying the cuts to both the leading and subleading photon candidates. The MVA methods used for event selection in the nominal analysis use specific SM MC training samples and most importantly one of the training variables used, namely  $\cos(\phi_1 - \phi_2)$ , is highly correlated to the angular variable,  $\cos(\theta_{\text{CS}}^*)$ , which can be used to distinguish spin hypotheses. Furthermore, given the unusual production modes of a spin-2 boson, no exclusive tagging is used in the spin analysis. The impact of using jet variables was studied but it was found that the sensitivity for distinguishing spin hypotheses was improved by a negligible amount.



**Figure 5.10:** The distribution of  $|\cos(\theta_{CS}^*)|$  before any selection cuts (left) and after the selection cuts (right) for the 7 TeV dataset (top row) and the 8 TeV dataset (bottom row). The three histograms represent the spin  $0^+$  distribution with all SM production modes (red circular points), the spin  $2_m^+$  distribution with the gluon-fusion production mode (blue square points) and the spin  $2_m^+$  distribution with the quark-antiquark annihilation production mode (green triangular points). The  $|\cos(\theta_{CS}^*)|$  category boundaries are shown as the black dashed lines.

The effect of the photon selection cuts on the distributions of  $|\cos(\theta_{CS}^*)|$  is illustrated in Fig. 5.10. Before any acceptance cuts, Fig. 5.10 (left), the  $|\cos(\theta_{CS}^*)|$  distribution of the  $0^+$  processes is flat. This is not the case for the  $2_m^+$  processes (gluon-fusion and quark-antiquark annihilation). After the selection cuts are applied these distributions are considerably distorted, Fig. 5.10 (right). As a Higgs produced from vector boson-fusion, which is  $\sim 8\%$  of the total (compared to  $\sim 88\%$  from gluon fusion), is typically produced at higher transverse momentum there is some additional contribution of  $0^+$  signal at high values of  $|\cos(\theta_{CS}^*)|$  compared to the  $2_m^+$  production modes after the selection cuts.

A robust analysis is possible because although the acceptance  $\times$  efficiency varies considerably as a function of  $|\cos(\theta_{CS}^*)|$ , the shape of this variation is largely independent of the spin-parity model. This is also true in restricted ranges of  $\eta$  and  $R_9$  which allows us to extract the signal yield in bins of  $|\cos(\theta_{CS}^*)|$  in a comparatively model-independent way.



**Figure 5.11:** Acceptance  $\times$  efficiency ratio between the  $2_m^+$  (gluon-fusion production) and  $0^+$  (all SM production modes) of the event selection as a function of  $|\cos(\theta_{CS}^*)|$  split into the  $|\eta|$  and  $R_9$  categories defined in Table 5.2. The  $|\cos(\theta_{CS}^*)|$  category boundaries are shown as the black dashed lines. The left hand plot is for the 7 TeV and the right hand plot for the 8 TeV.

**Table 5.2:** Definition of the photon resolution categories in the spin analysis. Here  $|\eta|_{\max}$  and  $R_{9\min}$  refer to the maximum  $\eta$  and minimum  $R_9$  of the two photons.

Category 0	$ \eta _{\max} < 1.444$ and $R_{9\min} \geq 0.94$
Category 1	$ \eta _{\max} < 1.444$ and $R_{9\min} < 0.94$
Category 2	$ \eta _{\max} > 1.556$ and $R_{9\min} \geq 0.94$
Category 3	$ \eta _{\max} > 1.556$ and $R_{9\min} < 0.94$

Figure 5.11 shows the efficiency  $\times$  acceptance ratio between the  $2_m^+$  (with gluon-fusion production only) and  $0^+$  (all SM production modes) as a function of  $|\cos(\theta_{CS}^*)|$  in the  $|\eta|$  and  $R_9$  categories defined in Table 5.2. It is clear that the acceptance  $\times$  efficiency between the spin-0 and spin-2 models is independent of  $|\cos(\theta_{CS}^*)|$  apart from at high values of  $|\cos(\theta_{CS}^*)|$  where the vector boson-fusion production in the SM plays a role. This motivates the choice of  $|\cos(\theta_{CS}^*)|$  category boundaries described below where all the categories have similar efficiency  $\times$  acceptance apart from the bin highest in  $|\cos(\theta_{CS}^*)|$ .

To benefit from the improved energy resolution of non-showering photons in the barrel, each event is categorised in  $\eta$  and  $R_9$  according to Table 5.2.

Within each category events are binned in  $|\cos(\theta_{CS}^*)|$ , to discriminate between the different spin hypotheses, according to Table. 5.3.

**Table 5.3:** Definition of diphoton  $|\cos(\theta_{\text{CS}}^*)|$  categories in the spin analysis.

Spin Category 0		$ \cos(\theta_{\text{CS}}^*)  < 0.2$
Spin Category 1	$0.2 \leq$	$ \cos(\theta_{\text{CS}}^*)  < 0.375$
Spin Category 2	$0.375 \leq$	$ \cos(\theta_{\text{CS}}^*)  < 0.55$
Spin Category 3	$0.55 \leq$	$ \cos(\theta_{\text{CS}}^*)  < 0.75$
Spin Category 4	$0.75 \leq$	$ \cos(\theta_{\text{CS}}^*)  < 1.0$

The  $|\cos(\theta_{\text{CS}}^*)|$  boundaries are optimised to make particular use of the most discriminating bin (high  $|\cos(\theta_{\text{CS}}^*)|$ ) and to maintain uniform acceptance  $\times$  efficiency in the other bins. In total the analysis is split into 20 event classes ( $4 \eta/R_9$  categories  $\times$  5  $|\cos(\theta_{\text{CS}}^*)|$  categories) in each year which gives a total of 40 event classes.

### 5.2.5 Event categorisation summary

All the categories and the tagging order are summarised in Table 5.4.

**Table 5.4:** The event classes at 7 and 8 TeV and some of their main selection requirements. Events are tested against the selection requirements of the classes in the order they are listed here. A breakdown of the amount of signal, background and data in each category is shown in Fig. 6.4.

Label	No. of classes		Main requirements
	7 TeV	8 TeV	
$t\bar{t}H$ lepton tag	★	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 1/2$ 1 b-tagged jet + 1 electron or muon diphoton BDT $> 0.6(-0.6)$ at 7(8) TeV
$VH$ tight $\ell$ tag	1	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 3/8$ $e$ or $\mu$ , $p_T > 20$ GeV, and $\cancel{E}_T > 45$ GeV OR $2e$ or $2\mu$ , $p_T > 10$ GeV; $70 < m_{ll} < 110$ GeV diphoton BDT $> 0.1(-0.6)$ at 7(8) TeV
$VH$ loose $\ell$ tag	1	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 3/8$ $e$ or $\mu$ , $p_T > 20$ GeV diphoton BDT $> 0.1(-0.6)$ at 7(8) TeV
$VBF$ dijet tag	2	3	$p_T^{\gamma_1}/m_{\gamma\gamma} > 1/2$ 2 jets; dijet and combined diphoton-dijet BDTs used
$VH$ $\cancel{E}_T$ tag	1	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 3/8$ $\cancel{E}_T > 70$ GeV diphoton BDT $> 0.8(0.0)$ at 7(8) TeV
$t\bar{t}H$ multijet tag	★	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 1/2$ 1 b-tagged jet + 4 more jets diphoton BDT $> 0.6(-0.2)$ at 7(8) TeV
$VH$ dijet tag	1	1	$p_T^{\gamma_1}/m_{\gamma\gamma} > 3/8$ jet pair, $p_T > 40$ GeV and $60 < m_{jj} < 120$ GeV diphoton BDT $> 0.6(0.2)$ at 7(8) TeV
Untagged	4/8/20 <sup>†</sup>	5/10/20 <sup>†</sup>	The remaining events, classified using diphoton BDT. These contain mostly $ggH$ production.

★ For the 7 TeV dataset, events in the  $t\bar{t}H$  lepton tag and multijet tag classes are combined, after selection, to form a single event class.

† For the MFM there are 4 (5) categories at 7 (8) TeV, for the SMVA there are 8 (10) and for the spin analysis there are 20 (20) with no exclusive categories.



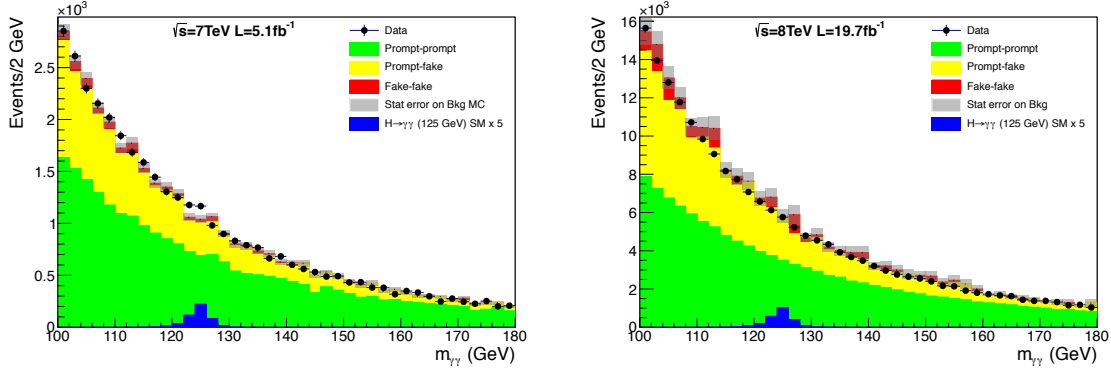
# Chapter 6

## Analysis

The analysis culminates in the search for a SM-like Higgs boson decaying into two photons in the mass range  $110 \leq m_H \leq 150$  GeV. This is done by making use of two variables which have very different shapes in signal and background. The first is the invariant mass of the two photons,  $m_{\gamma\gamma}$ , relative to the expected signal position,  $m_H$ , for which the signal is peaking and the background is smoothly falling. The second is the output of the diphoton BDT discussed in Sec. 5.1.3 in which the signal peaks towards positive values and the background peaks towards negative values. The distribution of the invariant mass is shown for data, background and signal MC simulation in Fig. 6.1 for all events which pass the MFM analysis cuts described in the previous chapter. The distribution of the diphoton BDT output is shown in the previous chapter in Fig. 5.4. In the SMVA analysis these two variables are combined and used to extract a number of analysis bins (see Sec. 5.2.3) in which events are counted. In the MFM events are coarsely categorised according to the diphoton BDT output and then parametrically fitted as a function of the invariant mass  $m_{\gamma\gamma}$ . The methods used for extracting the background and signal expectations for both analyses are described in this chapter. The important characteristics which define the signal are its yield relative to the SM expectation,  $\mu = \sigma/\sigma_{\text{SM}}$ , and its position,  $m_H$ .

### 6.1 Signal modelling

The signal MC samples (described in Sec. 4.2.1) are propagated through the full analysis separately for each production mechanism ( $ggH$ ,  $VBF$ ,  $WH$ ,  $ZH$ ,  $t\bar{t}H$ ). Events in the samples are weighted by the relevant SM cross section, branching ratio, the integrated



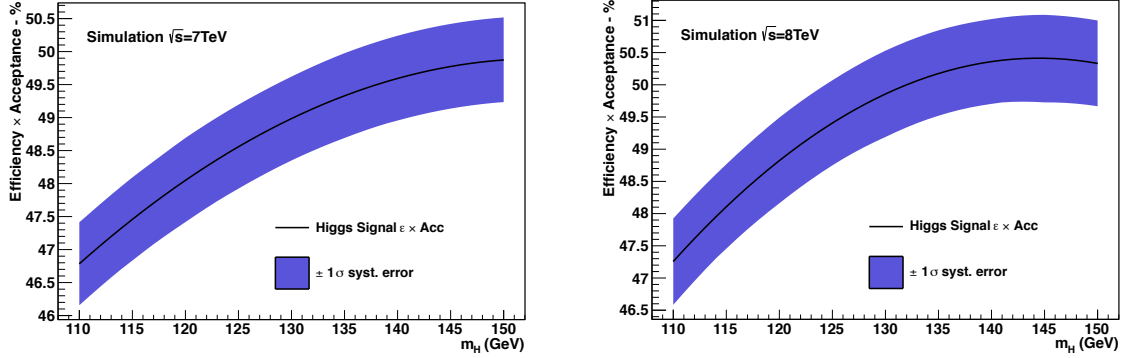
**Figure 6.1:** The diphoton invariant mass distribution for the 7 (left) and 8 TeV (right) datasets for the events which pass all of the analysis cuts. Cross-section weighted MC events are plotted as the filled histograms for the prompt-prompt (green), prompt-fake (yellow) and fake-fake (red) backgrounds. The SM signal expectation scaled up by a factor of 5 is shown by the blue histogram. The grey bands show the statistical uncertainty (sum of weights) of the background MC events.

luminosity and the “detector” weight which compensates for mismodelling in the MC such as pileup, the beamspot width and discrepancies between the efficiency in data and MC as measured in  $Z \rightarrow e^+e^-$  and  $Z \rightarrow \mu^+\mu^-\gamma$  decays. In this way the efficiency and acceptance of the detector, selection and categorisation is mapped by the MC samples and the total number of expected SM signal events is given by the sum of weights of the sample. The MC is generated separately for hypothesised Higgs masses,  $m_H$ , in the search range of  $110 \leq m_H \leq 150$  GeV in steps of 5 GeV. For any intermediate points the signal model is interpolated. The efficiency times acceptance,  $\epsilon \times \alpha$ , of the analysis selection for a SM Higgs boson is shown in Fig. 6.2 as a function of the Higgs boson mass,  $m_H$ . For a Higgs at  $m_H = 125$  GeV  $\epsilon \times \alpha = 48.6 \pm 0.7\%$  ( $49.4 \pm 0.7\%$ ) for the 7 (8) TeV datasets.

### 6.1.1 Mass-factorised analysis

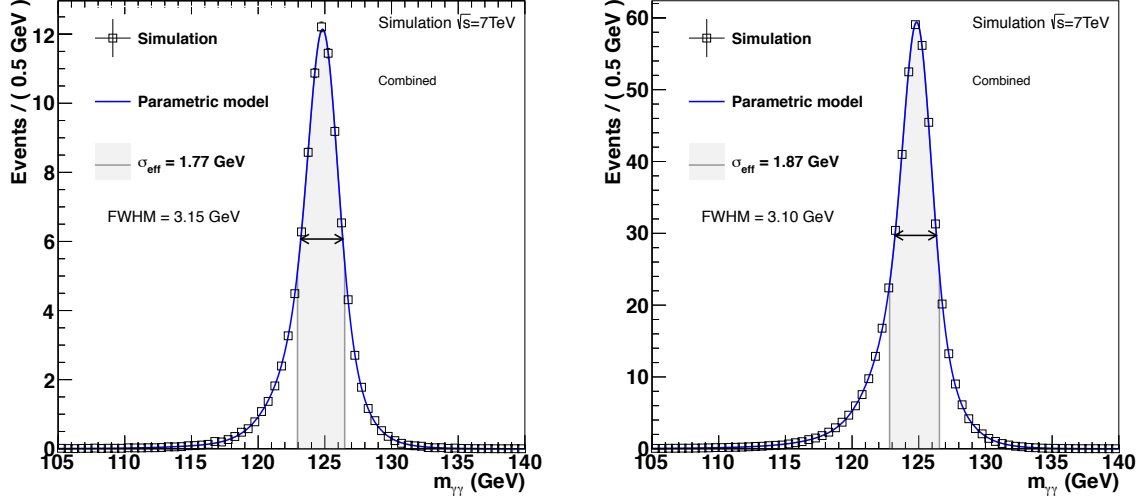
The diphoton invariant mass signal shape is modelled in a fully parametric way for the MFM. A separate model, consisting of a sum of Gaussians, is constructed for each production mechanism and each event class, where the two cases of right vertex and wrong (misidentified) vertex are fitted separately. The number of Gaussians used in the sum varies and can be as many as five although usually two or three provides an accurate description of the signal shape. A Gaussian sum is used because it has been found that this most accurately represents the shape of the invariant mass in signal. Many of the





**Figure 6.2:** The efficiency  $\times$  acceptance of the analysis selection for a SM Higgs signal as a function of Higgs mass for the 7 TeV (left) and 8 TeV datasets.

event classes contain a mixture of events from the resolution phase space (i.e. there are mixtures of barrel/endcap and converted/unconverted photons) so by fitting with a sum of Gaussians these different components, which are in themselves Gaussian or close to Gaussian distributed, are accounted for. Some of the event classes for particular production mechanisms contain rather low MC statistics (for example the lepton tagged category for signal produced by gluon fusion) and so fewer Gaussians are used in these cases. Clearly one expects a low signal yield for these particular cases and so a very accurate description of the shape is unnecessary. By allowing all the parameters in the Gaussian sums to float, including the means of the different Gaussian components, one obtains a parametrisation that models both the peak and tails of the signal distribution accurately. These fits are carried out at each value of  $m_H$  where signal MC is available and then the individual fit parameters are linearly interpolated such that the shape is defined for any  $m_H$  in the range  $[110, 150]$  GeV. The model therefore includes a floating parameter for the mass of the Higgs which can be fit to data. The right and wrong vertex shapes are then summed, with the relative fraction of right/wrong vertex events, to give a unique signal shape model in each event class for each production process. The signal normalisation, i.e. the expected number of SM signal events, is obtained by quadratically interpolating the efficiency  $\times$  acceptance for each event class and production process as calculated at the  $m_H$  points for which there is MC. The total signal shape, summed over event classes and production mechanisms, is shown in Fig. 6.3. The total number of expected signal events, as well as the  $\sigma_{\text{eff}}$  (minimum interval containing 68.3% of the distribution) and the  $\sigma_{\text{FW}}$  (full width at half the maximum divided by 2.35), for each event class are shown in Table 6.1. This information is additionally represented diagrammatically in Fig. 6.4 which provides an insightful summary of the analysis. The

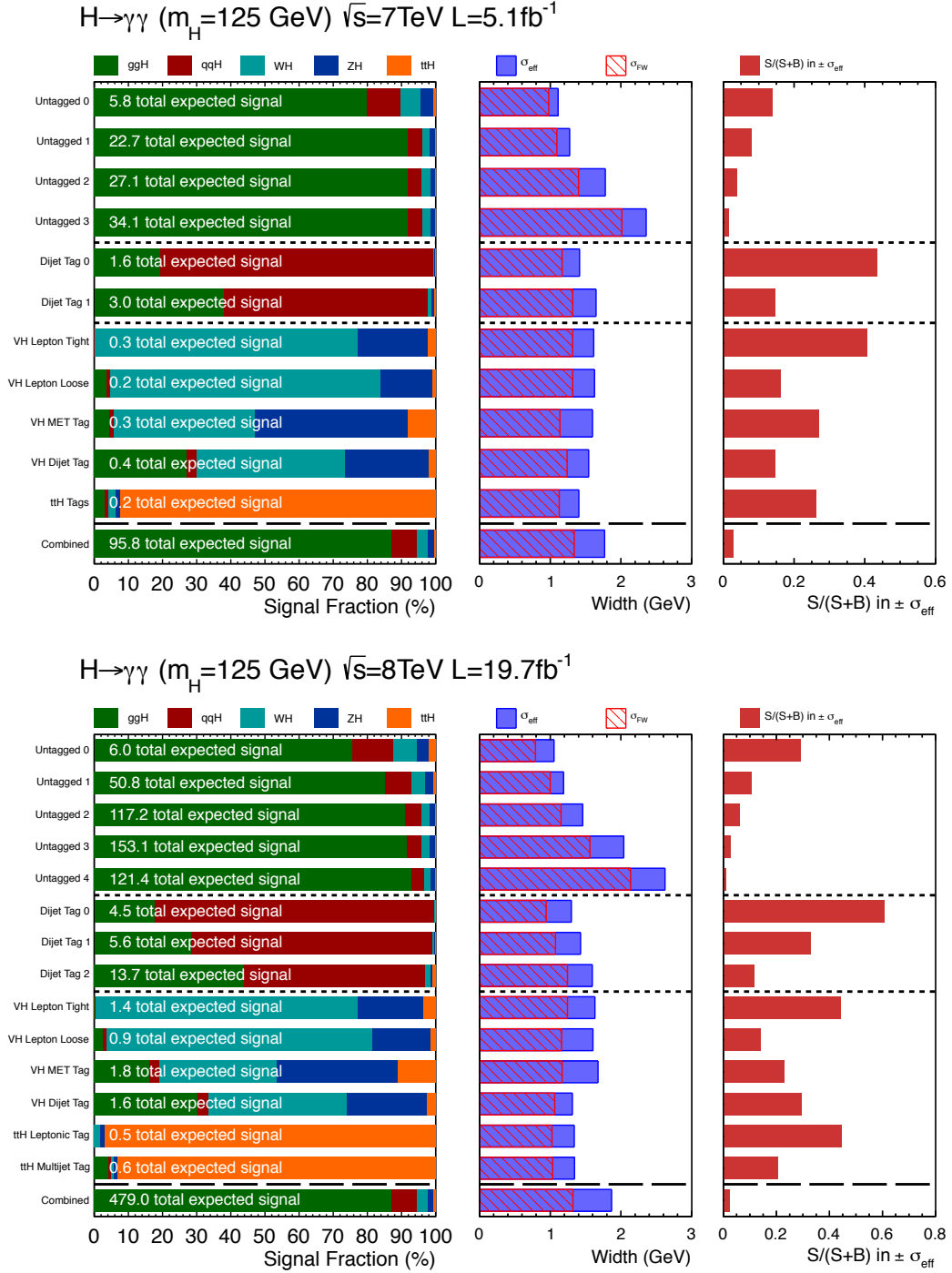


**Figure 6.3:** The diphoton invariant mass shape of the SM signal for the 7 TeV dataset (left) and 8 TeV dataset (right). The black square points show the distribution in MC events, where the sum of weights represents the expected number of signal events from a SM Higgs boson at 125 GeV. The blue line shows the shape of the parametric model used to represent the signal.

left hand column of this figure (Fig. 6.4) shows the breakdown of each signal process in each of the analysis categories. It can be seen that the *untagged* categories contain mostly gluon fusion (green band) where the contribution of the other signal processes decreases when moving from “Untagged 0” to “Untagged 4”. This is as we expect given that the signal processes which aren’t gluon fusion produce Higgs bosons at higher  $p_T$  and thus get a higher diphoton BDT score (see Fig. 5.4). A similar pattern can be seen for the dijet-tagged categories for which the signal consists of predominantly  $VBF$  (red) but increasing amounts of  $ggH$  (green) when moving from “Dijet 0” - “Dijet 2”. Similarly the  $VH$ -tagged categories consist of mainly  $WH$  and  $ZH$  signal (turquoise and blue) and the  $t\bar{t}H$ -tagged categories consists of mainly  $t\bar{t}H$  signal (orange). The middle column of this figure shows the signal model width (in terms of  $\sigma_{\text{eff}}$  and  $\sigma_{\text{FW}}$ ) and demonstrates that categories which relate to a high score in the diphoton BDT (“Untagged 0”/“Untagged 1”) and a high score in the combined dijet-diphoton BDT (“Dijet 0”) are those with the best mass resolution. The right hand column of this figure shows the  $S/(S+B)$  ratio under the peak (within  $\pm \sigma_{\text{eff}}$  of  $m_H$ ) and demonstrates the sensitivity of the individual categories.

**Table 6.1:** Expected number of SM Higgs boson events ( $m_H = 125$  GeV) and estimated background density (“Bkg.”) at  $m_{\gamma\gamma} = 125$  GeV for all event classes of the 7 and 8 TeV datasets. The composition of the SM Higgs boson signal in terms of the production processes and its mass resolution is also given. Numbers are omitted for production processes contributing less than 0.05% to the signal. Much of this information is pictorially represented in Fig. 6.4.

Event classes		SM Higgs boson expected signal ( $m_H = 125$ GeV)								Bkg. (GeV <sup>-1</sup> )
		Total	$ggH$	$VBF$	$WH$	$ZH$	$t\bar{t}H$	$\sigma_{\text{eff}}$ (GeV)	$\sigma_{\text{FW}}$ (GeV)	
7 TeV 5.1fb <sup>-1</sup>	Untagged 0	5.8	79.8%	9.9%	6.0%	3.5%	0.8%	1.11	0.98	11.0
	Untagged 1	22.7	91.9%	4.2%	2.4%	1.3%	0.2%	1.27	1.09	69.5
	Untagged 2	27.1	91.9%	4.1%	2.4%	1.4%	0.2%	1.78	1.40	134.6
	Untagged 3	34.1	92.1%	4.0%	2.4%	1.3%	0.2%	2.36	2.01	311.9
	VBF dijet 0	1.6	19.3%	80.1%	0.3%	0.2%	0.1%	1.41	1.17	0.5
	VBF dijet 1	3.0	38.1%	59.5%	1.2%	0.7%	0.4%	1.65	1.32	3.5
	VH tight $\ell$	0.3	—	—	77.2%	20.6%	2.2%	1.61	1.31	0.1
	VH loose $\ell$	0.2	3.6%	1.1%	79.1%	15.2%	1.0%	1.63	1.32	0.2
	VH $\cancel{E}_T$	0.3	4.5%	1.1%	41.5%	44.6%	8.2%	1.60	1.14	0.2
	VH dijet	0.4	27.1%	2.8%	43.7%	24.3%	2.1%	1.54	1.24	0.5
	$t\bar{t}H$ tag	0.2	3.1%	1.1%	2.2%	1.3%	92.3%	1.40	1.13	0.2
8 TeV 19.7fb <sup>-1</sup>	Untagged 0	6.0	75.7%	11.9%	6.9%	3.6%	1.9%	1.05	0.79	4.7
	Untagged 1	50.8	85.2%	7.9%	4.0%	2.4%	0.6%	1.19	1.00	119.6
	Untagged 2	117.2	91.1%	4.7%	2.5%	1.4%	0.3%	1.46	1.15	418.2
	Untagged 3	153.1	91.6%	4.4%	2.4%	1.4%	0.3%	2.04	1.56	870.3
	Untagged 4	121.4	93.1%	3.6%	2.0%	1.1%	0.2%	2.62	2.14	1401.3
	VBF dijet 0	4.5	17.8%	81.8%	0.2%	0.1%	0.1%	1.30	0.94	0.8
	VBF dijet 1	5.6	28.5%	70.5%	0.6%	0.2%	0.2%	1.43	1.07	2.7
	VBF dijet 2	13.7	43.8%	53.2%	1.4%	0.8%	0.8%	1.59	1.24	22.1
	VH tight $\ell$	1.4	0.2%	0.2%	76.9%	19.0%	3.7%	1.63	1.24	0.4
	VH loose $\ell$	0.9	2.6%	1.1%	77.9%	16.8%	1.5%	1.60	1.16	1.2
	VH $\cancel{E}_T$	1.8	16.3%	2.7%	34.4%	35.4%	11.1%	1.68	1.17	1.3
	VH dijet	1.6	30.3%	3.1%	40.6%	23.4%	2.6%	1.31	1.06	1.0
	$t\bar{t}H$ lepton	0.5	—	—	1.6%	1.6%	96.8%	1.34	1.03	0.2
	$t\bar{t}H$ multijet	0.6	4.1%	0.9%	0.8%	0.9%	93.3%	1.34	1.03	0.6



**Figure 6.4:** The expected composition and resolution of the signal for a SM Higgs at 125 GeV in the 7 TeV dataset (top) and 8 TeV dataset (bottom). The left diagram shows the breakdown of the signal in each event class by category:  $ggH$  (green),  $VBF$  (red),  $WH$  (turquoise),  $ZH$  (blue),  $t\bar{t}H$  (orange). The middle diagram shows the expected resolution in each event class in terms of  $\sigma_{\text{eff}}$  and  $\sigma_{\text{FW}}$ . The right diagram shows the expected signal over signal-plus-background ratio in a window of  $\pm \sigma_{\text{eff}}$  around  $m_H = 125$  GeV.

### 6.1.2 Sideband analysis

The statistical method used for the sideband analysis is not a parametric shape analysis like the MFM but implemented as a simple counting experiment, across the analysis categories, inside the signal region, which is defined as a  $\pm 2\%$  window around the hypothesised Higgs mass,  $m_H$ . The signal invariant mass shape is included by the categorisation scheme described in Sec. 5.2.3 whereby events nearer the signal peak are collected in bins with higher sensitivity (see Fig. 5.9). The number of expected signal events in each category is simply obtained from the sum of weights of the MC samples in each category which fall inside window for values of  $m_H$  for which there exists MC samples. The signal shape, i.e. the signal distribution across the analysis categories, is linearly interpolated for any intermediate values of  $m_H$ . As for the MFM the SM production mechanisms are propagated through the analysis separately. The signal normalisation at each Higgs mass is calculated in a similar way as for the MFM in which the  $\epsilon \times \alpha$  is linearly interpolated between the masses at which there are MC samples and then scaled by cross section, branching ratio and integrated luminosity.

### 6.1.3 Spin analysis

In the spin analysis the signal models are obtained from MC simulation as described in Sec. 4.2.1 for the spin-0 SM processes and the spin-2 processes. A parametric model identical to the one built for the nominal analysis is constructed as per the method described in Sec. 6.1.1. The acceptance  $\times$  efficiency of the two spin models in each category as well as the differential cross section as a function of  $|\cos(\theta_{cs}^*)|$ , which depends only on the spin of the initial state, are obtained from the MC simulation. The only remaining assumption is on the total number of expected signal events for a given spin-parity state and production mode. This is well defined for the spin-0 SM case and is obtained from the  $\sigma \times BR$  given by the LHC Higgs cross section working group in [25]. For the graviton-like  $2_m^+$  this quantity is unknown. Consequently we scale the signal models for both spin hypotheses with a modifier,  $\mu$ , such that when  $\mu = 1$  the total number of expected signal events for the model in question is equivalent to the SM expectation.

## 6.2 Background modelling

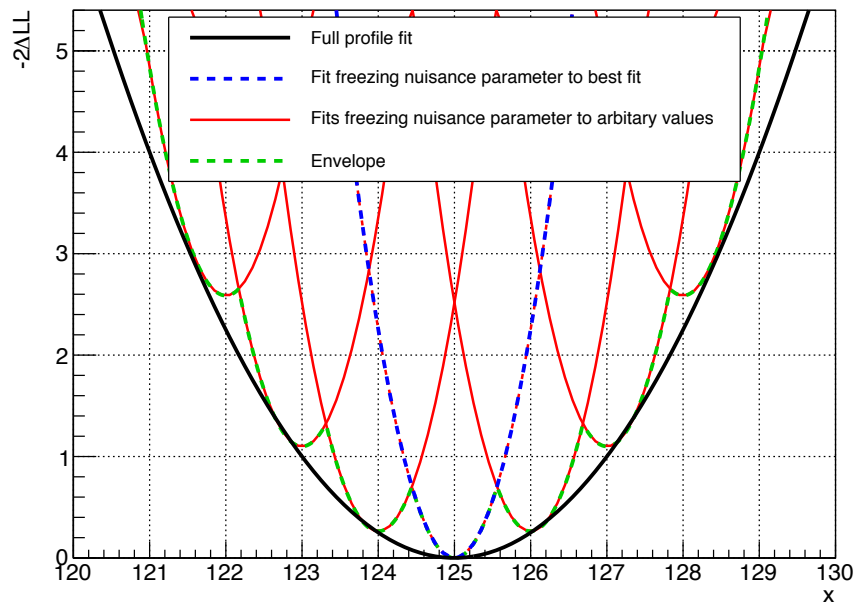
The background is the most significant unknown in this analysis. The size of the background relative to the signal is large and the invariant mass shape of the background is poorly modelled in MC. It is known to be a smoothly varying and falling spectrum, however various detector effects such as reconstruction, energy resolution and triggering which are imperfectly modelled in the MC can distort the shape. Furthermore the contribution of fake photons to the background varies as a function of  $m_{\gamma\gamma}$  and this is not well modelled in the theory or detector simulation. One of the main motivations for having the two analyses described here is that they have completely different ways of measuring the background and consequently serve to cross-check each other. Both analyses use entirely data-driven methods for extracting the background.

### 6.2.1 Mass-factorised analysis

A unique and novel way of estimating the background, given we have no *a priori* knowledge of its shape, has been developed specifically for this analysis. The method, latterly referred to as the “envelope” method, attempts to parametrise our ignorance of the background shape in a similar way to what is done for normal nuisance parameters when using the negative log likelihood to obtain the best fit value of a quantity and its error using frequentist statistical methods[69]. To explain better the method let us first consider the simplified case of fitting a probability distribution to a dataset where the probability distribution contains one physical parameter of interest (POI),  $x$ , (which could for example represent the signal mass) and one nuisance parameter,  $\theta$ , (which could represent the energy scale uncertainty). In this case we can calculate twice the negative log likelihood,  $-2LL$ , at different values of  $x$  whilst at each point minimising the likelihood with respect to  $\theta$ . This would give us the best fit value of  $x$  and its error given the variation allowed by  $\theta$ . This is represented in Fig. 6.5 by the solid black line. The best fit value of  $x$  is given at the minimum of the likelihood and the  $1\sigma$  error interval is defined as the range where twice the negative log likelihood relative to the global minimum (best fit value) is less than one,  $-2\Delta LL \leq 1$ . This is known as the “profile likelihood method” as the nuisance parameter,  $\theta$ , is “profiled” (i.e. floated) in the fit. In other words, for a given value of  $x$  the value of  $\theta$  which minimizes the likelihood is chosen. One can redo the likelihood scan fixing  $\theta$  to its value at the best fit and this gives a narrower likelihood curve, demonstrated by the dashed blue line in Fig. 6.5. In

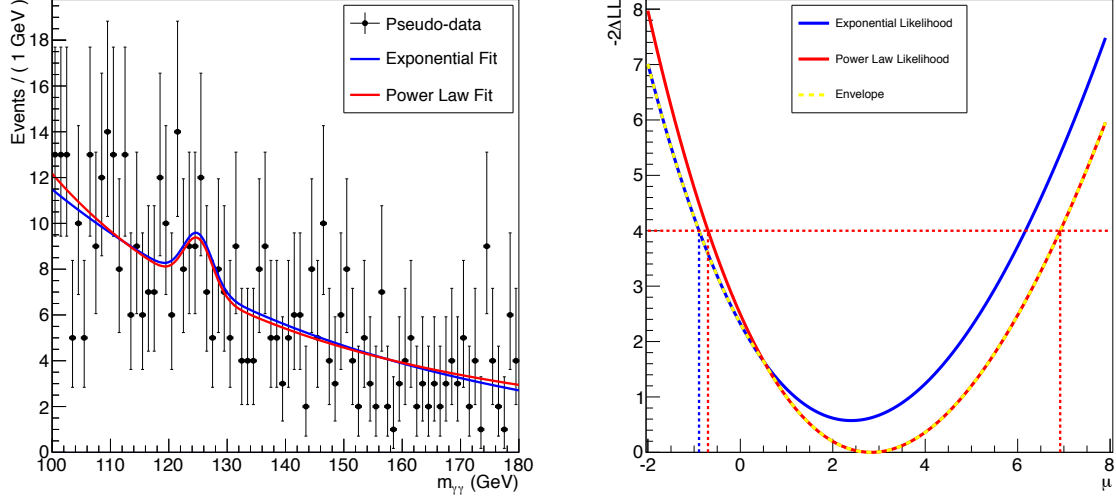
this case the error, the range for which  $-2\Delta\text{LL} \leq 1$ , is smaller as one would expect given the nuisance parameter is now frozen. This is equivalent to the statistical error only, as the systematic component parametrised by  $\theta$  is ignored. One can now also build up a multitude of curves by setting the nuisance parameter  $\theta$  to arbitrary values and rescanning the likelihood, these are shown in Fig. 6.5 as the red lines. It should be clear that by taking the “envelope” around the potentially infinite set of red curves one can reproduce the black line representing the full profile fit, providing the full  $\theta$  phase space is sampled enough times. This is shown in the figure by the green line which becomes smoother as more sets of  $\theta$  values are chosen. It is worth noting that not all of the red curves (of which there are infinitely many) necessarily have to touch the black line: a very extreme value of  $\theta$  will give a bad fit and the corresponding red dashed curve would be off the plot. In this way one could in principle “reverse-engineer” the profile likelihood method such that the profiling of the nuisance parameter  $\theta$  is not done as a continuous minimisation but as a series of discrete minimisations, where the minimum  $-2\Delta\text{LL}$  for a given value of  $x$  is taken as the minimum  $-2\Delta\text{LL}$  over all discrete choices of  $\theta$  at this value of  $x$ . Clearly, for a case in which the nuisance parameter  $\theta$  is continuous this is inefficient and unnecessary but it is effective if the nuisance parameter can only take discrete values.

In this analysis the background parametrisation is entirely unknown. So in principle if we could sample the infinite phase space of possible function choices we could use the method just described to profile the choice. In this way we find the function that minimises the negative likelihood for any given value of the parameter of interest  $x$ . This will then pick the function which fits the data best (minimises the negative likelihood) but can enlarge the error given that many different functions are tried. Let’s now consider a simplified case of the real situation. Imagine we have a large steeply falling background which can be parametrised by two possible choices: a single term power law,  $x^{-p}$ , and a single term exponential,  $e^{-px}$ . We have a small Gaussian-like signal component and our POI is the size of this signal,  $\mu$ . The best fit distributions are shown for a generated pseudo-dataset in the left hand plot of Fig. 6.6. The right hand plot of Fig. 6.6 shows the likelihood scan across the parameter  $\mu$ , which represents the size of the signal, for the two chosen functions (blue and red lines). The envelope (the minimum of the negative likelihood across both function choices) is shown as the yellow dashed line. One can see that the global best fit from the envelope is at a value of  $\mu = 2.83$  which is that obtained with the power law function. The  $1\sigma$  error (the point where the  $-2\Delta\text{LL}$  crosses 1) is unchanged in the envelope with respect to the result using the power law function alone. However, the  $2\sigma$  error (the point where the  $-2\Delta\text{LL}$  crosses 4) is increased with



**Figure 6.5:** Conceptual idea of the method of “profiling the likelihood”. Here the  $-2\Delta\text{LL}$  is scanned as a function of a physical parameter of interest  $x$  with a nuisance parameter  $\theta$  in three cases: 1) where  $\theta$  is freely floating (black line), 2) where  $\theta$  is set to its value at the  $-2\Delta\text{LL}$  minimum (blue line), 3) for several arbitrary choices of  $\theta$  (red line). The green line shows the “envelope” around the total minimum over several discrete choices of  $\theta$ . After an infinite sampling of  $\theta$  the green line would match the black.





**Figure 6.6:** An example of using the envelope in a realistic situation. The plot on the left shows a signal plus background fit to some pseudo-data using two different background function choices: a single term power law (red) and a single term exponential (blue). The plot on the right shows the profile likelihood curve for the signal size,  $\mu$ , for the two different function choices. The global minimum is at  $\mu = 2.83$  and the power law is used as the background choice. The envelope likelihood curve, profiling over the function choices is shown as the yellow dashed line.

respect to the result using the power law function alone. This is the principle behind the “envelope” background method. One can see that it is analogous to using a normal nuisance parameter which can only take discrete values. In this case the discrete values index which function is chosen. It means that a specific function choice never has to be made and the error on a given value will increase to account for situations where two or more functions give a similarly good fit. However, there is one more important feature of the method which must be discussed before it can be applied to the data.

The concept of the “envelope” method has now been demonstrated with a simplified example in which two function choices were profiled and both these functions had one free parameter. However we would like to be able to sample as much of the function phase space possible (i.e. as many functions as we can) which means there will be choices with different numbers of free parameters. Two functions of the same type but of a different order may give a very similar looking fit but functions with more free parameters are more flexible and thus will give a lower negative likelihood value. For example consider the functions  $E_1 = e^{-p_1 x}$  and  $E_2 = f_1 e^{-p_1 x} + (1 - f_1) e^{-p_2 x}$ , which are first and second order exponential sums and have one and three free parameters respectively, not including an overall normalisation term. There may be a case where these give an identical fit (as

$f_1 \rightarrow 1$  or  $p_1 \rightarrow p_2$ ) however the negative log likelihood will always be lower for a higher parameter fit and therefore the higher order function would always be the minimum of the envelope. This means that for any embedded class of functions, the function which gives the global minimum of the likelihood will be of the highest order allowed. Consequently, a correction scheme has been devised to avoid this problem and to penalise functions in the envelope which have more free parameters but don't necessarily fit the data any better.

The negative log likelihood gets redefined as:

$$-2LL = -2\ln(\mathcal{L}) + cN_p, \quad (6.1)$$

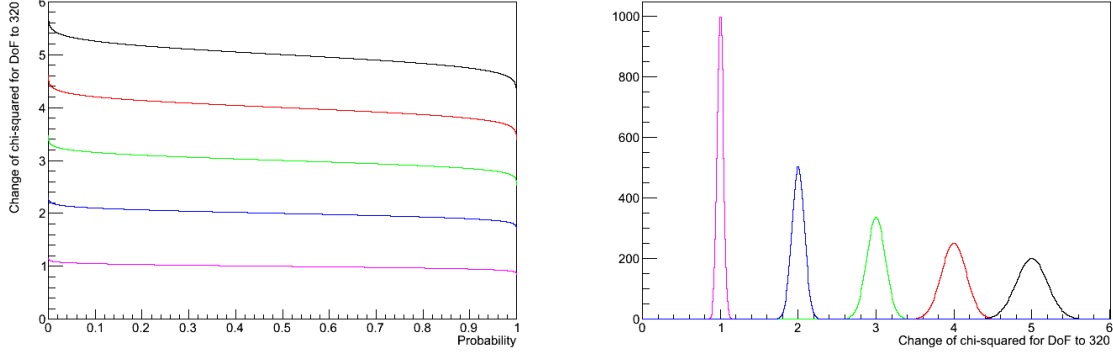
where  $\mathcal{L}$  is the likelihood for the function,  $N_p$  is the number of free parameters in that function and  $c$  is a constant correction term. Two correction schemes were studied, for values of  $c = 1$  and  $c = 2$ . The motivation for choosing  $c = 2$  is an application of the Akaike information criterion, described in Ref.[70], which states that for large sample sizes the corrected likelihood is

$$A = -2\ln(\mathcal{L}) + 2N_p. \quad (6.2)$$

It can be seen that the correction term here is  $2N_p$ , which gives a value of  $c = 2$ , in others words a correction of 2 per free parameter.

The argument for using a correction of  $c = 1$  is a little more natural and arises from the assumption that in the high statistics limit for a binned dataset,  $-2\Delta LL \approx \chi^2$ . Given a value of the  $\chi^2$  one can calculate the  $\chi^2$   $p$ -value given the number of degrees of freedom,  $N$ , which is equal to the number of bins,  $b$ , minus the number of free parameters in the fit function,  $N_p$ . Thus the  $p$ -value can be expressed as  $p(\chi^2, b - N_p)$ . One can then determine a new chi-squared value,  $\chi'^2$ , defined as the one which would give the equivalent  $p$ -value but with a different number of degrees of freedom, namely the one in which there are no free parameters in the fit,  $p(\chi'^2, b)$ . It can be seen that there is now an expression for  $\chi'^2$ , which is approximately equivalent to a new  $-2\Delta LL$  (now relative to the best possible fit, given the data), which is independent of the number of fit parameters,  $N_p$ , and thus the corrected likelihood is given by,

$$-2LL = -2\ln(\mathcal{L}) + \chi'^2 - \chi^2. \quad (6.3)$$



**Figure 6.7:** The value of the correction,  $\chi'^2 - \chi^2$ , as a function of the fit  $p$ -value for a fit with 320 bins is shown on the left. The projection of the correction integrated uniformly over  $p$ -values is shown on the right. The five different coloured lines represented fit functions with different numbers of free parameters, ranging from one free parameter (pink) up to five free parameters (black).

The correction term,  $\chi'^2 - \chi^2$ , depends on the number of bins, the number of free fit parameters and the quality of the original fit (the  $\chi^2$   $p$ -value). Figure 6.7 shows how the size of the correction varies for functions with different numbers of fit parameters. It can be seen that on average the correction is given by,

$$\chi'^2 - \chi^2 \approx N_p. \quad (6.4)$$

The correction scheme used in the analysis was chosen to be  $c = 1$ . This was decided using empirical results of studying the impact on the bias and error coverage when fitting the physical parameters of the signal yield,  $\mu$ , and the signal position,  $m_H$ , using the envelope method with either correction scheme. Before we cover the results and conclusions of this study, we will first discuss how one decides which functions to include in the envelope.

In principle it would be beneficial to choose any and every function one can think of that can reasonably describe a falling spectrum. In practice this is unfeasible as the combinatorics of the problem rapidly spiral out of control. If an analysis has multiple categories,  $N_c$ , and one chooses multiple background functions in each category,  $N_f$ , then the number of combinations goes like  $N_f^{N_c}$  which rapidly makes the problem computationally impossible. Instead, one has to choose a smaller number of functions which reasonably cover the phase space of infinite functions. The functions used in this analysis come in four main classes. They are as follows (note an overall normalisation term is not included in the equations below),

- *Exponential recursive sum* - a sum of terms like  $e^{-px}$ .

$$p(x) = c_1 e^{-p_1 x} + (1 - c_1) c_2 e^{-p_2 x} + (1 - c_2) c_3 e^{-p_3 x} + \dots + (1 - c_{n-1}) c_n e^{-p_n x} \quad (6.5)$$

- This has  $2n - 1$  free parameters per order. The functions are labelled by the number of free parameters so the lowest order is **exp1** then **exp3**, **exp5** and so on.

- *Power law recursive sum* - a sum of terms like  $x^{-p}$ .

$$p(x) = c_1 x^{-p_1} + (1 - c_1) c_2 x^{-p_2} + (1 - c_2) c_3 x^{-p_3} + \dots + (1 - c_{n-1}) c_n x^{-p_n} \quad (6.6)$$

- This has  $2n - 1$  free parameters per order. The functions are labelled by the number of free parameters so the lowest order is **pow1** then **pow3**, **pow5** and so on.

- *Laurent series* - The best fit value for a single order power term is around -4.3. Consequently we use a Laurent-like series of a sum of terms like  $x^{-n}$  expanded around  $x^{-4}$ .

$$p(x) = c_1 x^{-4} + (1 - c_1) c_2 x^{-5} + (1 - c_2) c_3 x^{-3} + (1 - c_3) c_4 x^{-6} + \dots \quad (6.7)$$

- This has  $n - 1$  free parameters per order. The functions are labelled by the number of free parameters so the lowest order is **lau1** then **lau2**, **lau3** and so on.

- *Bernstein polynomials* - These are polynomials in the Bernstein basis[71, 72]. A Bernstein polynomial of degree  $n$  is given by,

$$p(x) = \sum_{i=1}^n c_i \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i} \quad (6.8)$$

- This has  $n$  free parameters per order. The functions are labelled by the number of free parameters so the lowest order is **pol1** then **pol2**, **pol3** and so on.

One then has to determine which order functions of each of these classes is used in the envelope for a given dataset. To pick the lowest order included in the envelope a simple goodness of fit test is used and has the loose requirement that the  $\chi^2$   $p$ -value of the fit is greater than 0.01. To pick the highest order used in the envelope a Fisher test ( $f$ -test) requirement is imposed[73], which can be described as follows. If two functions of the

same class have  $n$  and  $n + m$  free parameters respectively then, in the high statistics limit, the difference in  $-2\Delta\text{LL}$  between the two functions is distributed as  $\chi^2$  with  $m$  degrees of freedom. Consequently the difference in  $-2\Delta\text{LL}$  between the two fits is converted into a  $\chi^2$   $p$ -value and the next higher order function is included if  $p < 0.1$ . In other words if the higher order function does not improve the fit sufficiently it is not included. This determines which functions are included in the envelope for each category and ranges anywhere between four and nine. It should be noted that the requirements to include a function are intentionally loose so that as much of the “function-space” is sampled as possible.

In order to assess the bias and coverage properties of the envelope method a few functions are chosen as “truth” models from which to generate pseudo-data and test the statistical validity of the method. A single function of each class is chosen as a “truth” model and the order determined by the same  $f$ -test but with a tighter requirement that  $p < 0.05$ . These models are first fit to the data and then the pseudo-data generated from these values. The comparison of the systematic bias to statistical uncertainty is determined by the pull distribution,

$$\text{pull}(\mu) = \frac{\mu_{fit} - \mu_{inj}}{\sigma_{fit}}, \quad (6.9)$$

where  $\mu_{inj}$  is the injected signal strength in the toy,  $\mu_{fit}$  is the fitted signal strength per toy and  $\sigma_{fit}$  is the error on the fitted signal strength per toy. The systematic bias to statistical comparison is then the deviation of the mean of the distribution from zero as compared to the width of this distribution. An unbiased method will give a mean close to 0 and a method which covers accurately will give a width equal to 1. The  $X\sigma$  coverage is defined as the fraction of toys for which  $\mu_{fit} - X\sigma_{fit} \leq \mu_{inj} \leq \mu_{fit} + X\sigma_{fit}$ .

When using a correction factor of  $c = 1$ , the systematic bias on the signal strength is less than 14% of the statistical error from the fit and the coverage is accurate for 0.5, 1, 2 and  $3\sigma$  (all the points tested). This was found to be true when generating toy experiments for a range of different signal strengths and for signal at different mass values. The bias and coverage have also been tested when removing the “truth” model from the set of envelope functions, when removing all functions of the same class as the “truth” model and when using various convoluted truth models such as a histogram spline of the data[74], a kernel density estimator (sum of Gaussians)[75] and a hybrid of different functions patched together. All of these cases also demonstrated a reasonable level of bias and coverage.

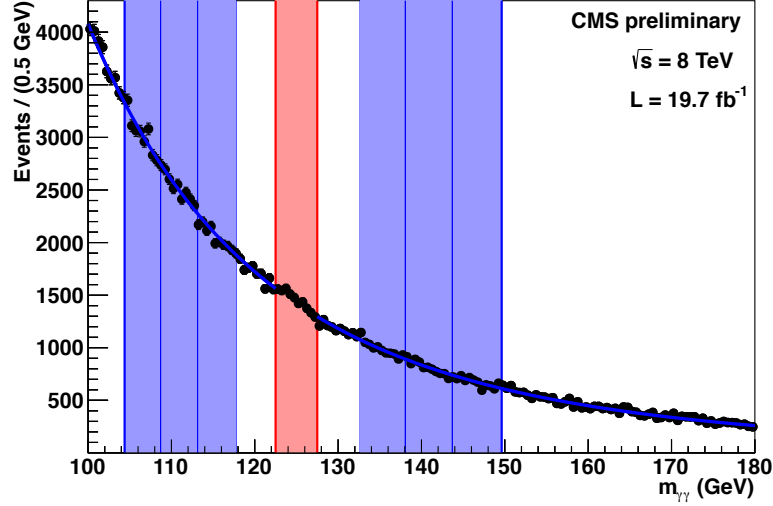
When using a correction factor of  $c = 2$  the systematic bias is larger than this. In extreme cases it can reach 30% of the statistical uncertainty and in these cases it is common that the method undercovers especially at higher standard deviations,  $\sigma$ . Consequently, for this analysis a correction factor of  $c = 1$  is chosen. The tests described above to ascertain the statistical validity of the method are performed for each individual category of the analysis and furthermore for the categories combined together in years and the whole ensemble of categories. The method has sensible behaviour in all of these cases.

The invariant mass distributions with the different envelope functions, after having been fit to the data, for each of the analysis categories are shown in Appendix A.

### 6.2.2 Sideband analysis

It is desirable, for the same reasons as in the MFM analysis, to extract the background estimation in the SMVA analysis from data. Given that the SMVA is simply a cut and count analysis in a signal window, of size  $\pm 2\%$ , around the hypothesised Higgs mass,  $m_H$ , a fully parametric description of the background in each category is not required. It is simply a case of estimating the number of expected background events in the signal region for each of the SMVA categories, which are defined by picking out regions of the two-dimensional phase space,  $\Delta m/m_H$  vs. diphoton BDT output, as shown in Fig. 5.9 and described previously in Sec. 5.2.3. The total expected number of background events and the fraction populating each of the SMVA categories are extracted separately.

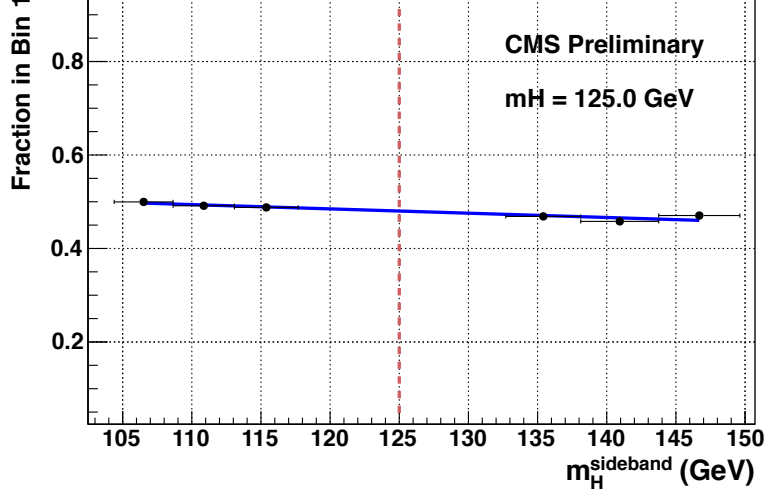
The total number of expected background events, i.e. the overall normalisation of the background, in the signal region is obtained by fitting the invariant mass distribution, summed over all the categories, in the range  $100 \leq m_{\gamma\gamma} \leq 180$  GeV where the data in the signal region is excluded from the fit. Figure 6.8 shows the invariant mass distribution for all events in data and the parametric fit used to extrapolate the number of background events in the signal region. The parametric form used to obtain the normalisation of the background is a single term power law (one degree of freedom) for the 8 TeV dataset and a two term Laurent series (one degree of freedom) for the 7 TeV dataset. The functions are defined identically to those described for the envelope method in Sec. 6.2.1. These functions are chosen because they give the smallest uncertainty when accounting for biases incurred by picking the wrong functional form. A systematic uncertainty is applied to the normalisation term of which there are more details in Sec. 6.3.



**Figure 6.8:** The inclusive invariant mass distribution for the 8 TeV dataset in data (black points). An illustration of the signal region (red) and the sidebands (blue) used for the mass hypothesis  $m_H = 125$  GeV is shown. The parametric function used to obtain the normalisation in the signal region is shown as the blue line.

The fraction of background events which populate each category is extracted using the data in invariant mass sidebands. The signal region is defined as  $\pm 2\%$  window around the hypothesised mass,  $m_H$ . Each sideband is defined to have the equivalent width of  $\pm 2\%$  relative to the invariant mass at the centre of the sideband, thus the sidebands on the upper side are wider than the sidebands on the lower side. For a SM Higgs at 125 GeV about 75% of the signal is contained within the window. Consequently, one sideband either side of the signal window is skipped and then three sidebands on either side of the hypothesised mass in question are used. Also shown in Fig. 6.8 are the signal region and the six sidebands for  $m_H = 125$  GeV. The Higgs masses tested in the sideband analysis are between  $115 \leq m_H \leq 150$  and consequently both the signal region and the sidebands slide given the value of  $m_H$  under consideration. Because of Drell-Yan contamination ( $Z \rightarrow e^+e^-$  misidentified as two photons) in the region of low invariant mass,  $m_{\gamma\gamma} < 100$  GeV, the sideband analysis only considers signal in the range  $115 \leq m_H \leq 150$  so that the bottom edge of the lowest sideband at  $m_H = 115$  GeV is above 100 GeV.

The data is then split into categories and a straight line fit is performed in each which uses the fraction of data events, contained in the given category, in each sideband to estimate the fraction of background events, contained in the given category, in the signal region. An example of one of these fits is shown in Fig. 6.9 for “Category 0” in the 8 TeV



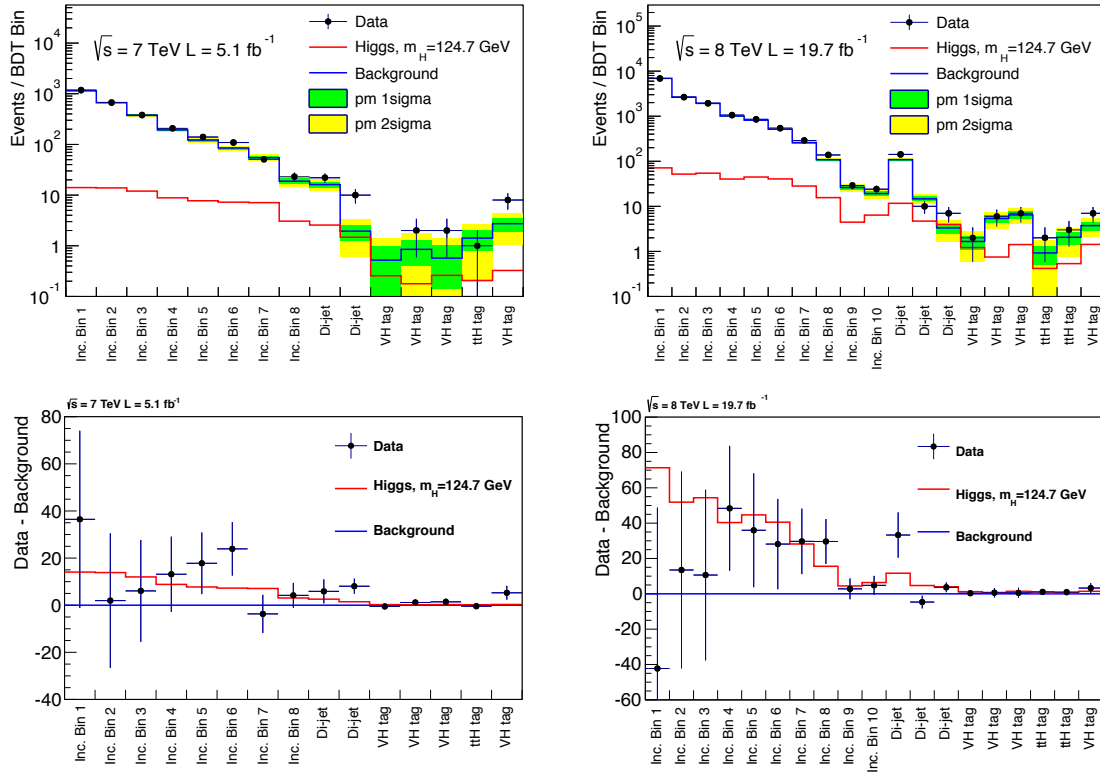
**Figure 6.9:** The fraction of data events in a single analysis category in each of the sidebands surrounding the signal window at  $m_H = 125$  GeV (black points), shown for the SMVA “Category 1” (or “Bin 1”) in the 8 TeV dataset. The blue line shows the straight line fit for this bin. In the 8 TeV dataset there are 10 inclusive analysis bins and 9 exclusive analysis bins. Consequently there are 19 of these straight line fits performed simultaneously at each  $m_H$  but only 18 of them are independent as the fraction of events obtained in the signal region must sum to unity.

dataset. These straight line fits are performed for each category simultaneously with the constraint that the sum across all the analysis categories of the fraction of background events estimated in the signal region is equal to one. It is assumed that the fraction of events in each category varies linearly with invariant mass (i.e. hence the straight line across sidebands) and that there is negligible signal contamination in the sidebands. For each category,  $c$ , the fraction of events for a given mass is taken to be

$$f_c = p_{0c} + p_{1c}(m - m_H), \quad (6.10)$$

where  $p_{0c}$  and  $p_{1c}$  are the straight line fit parameters for category  $c$ . Since the fractions must sum to unity for any given mass (the normalisation is extracted elsewhere) then for  $N$  bins there are  $2(N - 1)$  coefficients. These coefficients are determined by the straight line across the sidebands in each category, with the constraint that the fraction in each sideband must sum to one across all the categories. The distribution for the data, background (along with the  $\pm 1\sigma$  and  $\pm 2\sigma$  errors) and signal for each of the sideband analysis bins are shown in Fig. 6.10 for the mass hypothesis,  $m_H = 124.7$  GeV.





**Figure 6.10:** The distribution in sideband analysis categories (bins) for the data (black points), the background (blue line), the error on the background (green and yellow bands) and the signal (red line) for the mass hypothesis at  $m_H = 124.7$  GeV. The top row shows the expected number of events of each type for the 7 TeV (left) and 8 TeV (right) datasets. The bottom row shows the background subtracted plot for data (black points) and signal (red line).

### 6.2.3 Spin analysis

Two statistical tests are carried out in the spin analysis:

1. The signal strength,  $\mu$ , is extracted differentially in bins of  $|\cos(\theta_{\text{CS}}^*)|$ . This is a relatively model-independent test and in principle allows any spin model to be compared to the data.
2. The statistical separation between different spin hypotheses is calculated using a test statistic similar to the one described in Sec. 6.4 and the  $CL_s$  exclusion method is used to quantify the separation power. This is a highly model-dependent test but allows for the exclusion of specific spin models.

The background model for the spin analysis comes in two forms. For the differential measurement of the signal strength in bins of  $|\cos(\theta_{\text{CS}}^*)|$  the envelope background method is used as per the description in Sec. 6.2.1. However, when calculating the statistical separation between various spin hypotheses a single parametrisation of the background is used in each category, namely a polynomial in the Bernstein basis[71, 72] as per the description in Eq. 6.8. The reason for this is that there is no asymptotic approximation for the test statistic distribution when the null hypothesis is not embedded in the alternative hypothesis. Consequently, in order to obtain the test statistic distributions (like the ones shown in Fig. 6.11) one has to generate lots of pseudo-data and then refit this data to obtain the likelihood ratio and hence test statistic value. Given the complexity of the spin signal model and the combinatorics involved when using the envelope method with 40 analysis categories this becomes impractical in terms of computational power. It has been trialled using the GRID computing network but was found to take the order of hundreds of CPU years. Given this complication, and the fact that small losses in sensitivity to the background normalisation have a small impact on the spin hypothesis separation power, a single parametrisation in each category was chosen for ease and simplicity. The choice is to use 4th order Bernstein polynomials in all categories apart from the highest  $|\cos(\theta_{\text{CS}}^*)|$  categories in which a 3rd order was chosen. The motivation behind this choice is that these order of polynomials show a similarly small level of bias as the envelope method when tested against “truth” models for the spin categories analogous to the description in Sec. 6.2.1.

## 6.3 Systematic Uncertainties

There are several sources of systematic uncertainty in the analysis and these are described in the following section. Nearly all of these affect the signal model, although there is an uncertainty associated with the “envelope” background method, discussed in Sec. 6.2.1, for the MFM and there are two uncertainties to the background for the SMVA, discussed separately below. There is one additional systematic for the spin analysis, also discussed separately below. The signal systematics come in three broad categories; those associated with individual photons, those associated with individual events and those associated with specific tagged event classes. The systematics get applied to the model in different ways depending on the level of correlation required between them. The MFM analysis has a more precise correlation model between photon energy uncertainties and the mass shape, predominantly because this is used for the main result and is required to provide a measurement of the observed boson’s mass. For the SMVA this is somewhat simplified as the analysis only serves as a cross-check of the categorisation and background estimation for which the important systematics are on the overall signal yield and not so much its shape or position. The way the systematics are implemented in the two analyses is explained below, after which there is a description of the individual uncertainties used. There is a full summary of the systematics given in Table 6.2. The main groups of systematic uncertainty affecting the signal strength,  $\sigma/\sigma_{\text{SM}}$ , are shown in Table 6.3 and the main groups of systematic uncertainty affecting the uncertainty on the mass measurement,  $m_H$ , are shown in Table 6.4 at the end of the section.

### 6.3.1 Implementation of the systematic uncertainties

#### Mass-Factorised Analysis

In the MFM uncertainties are implemented in two different ways. The first type are generally associated with the photon energy scale and resolution and the differences between photons and electrons. These are implemented as nuisance parameters with a Gaussian constraint, and affect the position, shape and normalisation of the signal probability distribution function. The second type are uncertainties which result in events being misclassified or rejected by the event selection. These are implemented as

nuisance parameters with a log normal constraint of the following form:

$$f(x, \mu, \sigma) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad (6.11)$$

where  $\mu$  and  $\sigma$  represent the mean and standard deviation of the variable  $x$ 's natural logarithm, and affect the normalisation of the signal and the relative signal yield in each event class. Log normal constraints are chosen for the latter because they represent Gaussian-like constraints but are distorted such that the nuisance term cannot drive the yield negative. These uncertainties are somewhat easier to understand as they simply dictate that events can migrate between categories and that the overall normalisation of the signal can vary due to the theoretical cross section, luminosity measurement, etc. The former type is perhaps slightly harder to understand as these uncertainties are applied to individual photons and their effect is propagated through to the invariant mass shape and normalisation of the signal for each production process and event class. The methodology for this is as follows:

- Each uncertainty is designed to address a specific class of photons with similar properties. For example one uncertainty might be the energy scale uncertainty for unconverted photons in the barrel. We can label this class  $c$  and the uncertainty for this class  $\theta_c$  which is the nuisance parameter that enters the signal model.
- Within the class  $c$  there may be several subgroups of photons which have different uncertainties. For example photons in the central barrel and outer barrel. We will label this uncertainty as  $\sigma_s$ .
- The full analysis is run through where each photon in class  $c$  has its scale altered according to a random Gaussian number centred on its nominal value with a width of its individual uncertainty  $\sigma_s$  and all photons in the other classes are unchanged.
- The distributions of the diphoton invariant mass for the signal, with and without the above random shift, are compared. The relative difference between the mean, effective width and overall rate of these two distributions is calculated, labelled  $k_\mu$ ,  $k_\sigma$  and  $k_r$  respectively, for each signal process and analysis event class separately. This then provides a measure of the  $1\sigma$  systematic effect on the signal mass shape of the uncertainty  $\theta_c$ .

- Given there are several nuisance parameters  $\theta$  for which this is done, the signal model Gaussian parameters then get reparametrised as follows;

$$\text{Gaussian mean: } \mu(m_H) = \hat{\mu}(m_H) \left( 1 + \sum_c k_{\mu c} \theta_c \right) \quad (6.12)$$

$$\text{Gaussian sigma: } \sigma(m_H) = \hat{\sigma}(m_H) \left( 1 + \sqrt{\sum_c k_{\sigma c}^2 \theta_c^2} \right) \quad (6.13)$$

$$\text{Signal rate: } r(m_H) = \sigma_{\text{xs}} \cdot \mathcal{B} \cdot \epsilon \cdot \alpha \cdot L \left( 1 + \sum_c k_{rc} \theta_c \right) \quad (6.14)$$

where  $\hat{\mu}$  and  $\hat{\sigma}$  represent the nominal fitted values of the Gaussian means and widths in the MC and  $\sigma_{\text{xs}}$ ,  $\mathcal{B}$  and  $L$  represent the cross-section, branching ratio and luminosity respectively.

The final event classes contain mixtures of photons from several different regions (in  $\eta$  and  $R_9$ ) which all have different associated uncertainties. Using this method the full correlation between the individual photon uncertainties is completely mapped onto the final invariant mass signal shape of each analysis category. Each shape parameter of the signal model is given a dependence on each of the nuisance parameters. The signal shape can then be altered according to the size of the effect the nuisance parameter has in that region of signal phase space.

### Sideband analysis

As the sideband analysis is simply a cut-and-count experiment across 15 (19) bins in the 7 (8) TeV datasets the uncertainties are just implemented as normalisation errors affecting the yield in each bin. These can arise from the energy scale, resolution, reconstruction efficiency, etc., and result in event migrations between bins or arise from the theoretical cross section, luminosity measurement, etc., and affect the overall signal normalisation. There are two systematics in the SMVA associated to the background estimation from the sidebands. The first arises from the assumption that the background varies linearly across sidebands, although this has a negligible impact on the overall signal yield error. The second arises from the assumed shape used to extract the background normalisation. This is the dominant systematic in the analysis and is extracted using a bias study similar to the one described for the envelope method in Sec. 6.2.1. The size of this uncertainty is 0.7% (0.4%) on the background rate in the 7 (8) TeV dataset.

## Spin analysis

One additional uncertainty is used in the spin analysis to account for mismodelling of the  $p_T$  spectrum of the signal in MC events. There is a considerable correlation between the graviton  $p_T$  and the decay angle,  $\cos(\theta_{cs}^*)$ , and this distribution is not well known theoretically. Consequently, the graviton  $p_T$  spectrum is reweighted to match the SM prediction and an uncertainty is applied to allow events to migrate between the  $|\cos(\theta_{cs}^*)|$  categories given a 10% change in the graviton  $p_T$ .

### 6.3.2 Systematic uncertainties related to individual photons

#### Photon energy scale

In Section 4.3 a method for correcting the photon energy using  $Z \rightarrow e^+e^-$  decays was described. Although the statistical uncertainty on these corrections is small, there are some data/MC discrepancies that gives rise to a systematic uncertainty. This uncertainty is individually calculated for the 8 categories (4 in  $\eta \times 2$  in  $R_\eta$ ) for which the energy scale correction is applied. There are four nuisance parameters for each of the 7 and 8 TeV datasets propagated to the signal model representing the energy scale uncertainty for barrel/endcap and converted/unconverted photons.

#### Photon energy resolution

Section 4.3 also describes the method used to smear the MC events such that the energy resolution matches that observed in data using the  $Z \rightarrow e^+e^-$  control sample. As above there are four nuisance parameters representing the constant smearing uncertainty for each of the 7 and 8 TeV datasets for barrel/endcap and converted/unconverted photons. For converted/unconverted photons in the barrel there is an additional pair of uncertainties for the stochastic smearing term. Given the low sensitivity of the endcap region and the low statistics in the 7 TeV sample this additional stochastic parameterisation is only used in the barrel at 8 TeV.

### Energy scale uncertainty due to differences between electron and photon reconstruction

As the uncertainties, and the corresponding corrections, to the energy scale and resolution described above are derived from the  $Z \rightarrow e^+e^-$  decays, an important source of systematic uncertainty arises from those differences between electrons and photons which are not well modelled by the MC simulation. The important effect is not the absolute difference between electrons and photons, which is clearly different, but the differences between them in data which are not modelled accurately by the simulation.

The effect which is most significant is due to an imperfect description of the tracker and services material between the beampipe and the ECAL in the simulation. Studies of photon conversions, electron bremsstrahlung and pion scattering suggest a material deficit in the simulation of up to 20% in some areas upstream of the ECAL. The uncertainty on the energy scale is assessed by using specialised MC samples in which the tracking material budget is increased by 10% (in the barrel) and 20% (in the endcap) in 8 categories (4 in  $\eta \times 2$  in  $R_\eta$ ). Two nuisance parameters, which get correlated across the 7 and 8 TeV datasets, are implemented representing the energy scale uncertainty due to the material mismodelling in the central barrel,  $|\eta| < 1$ , and the rest of the ECAL,  $|\eta| \geq 1$ .

A further difference between electrons and photons which is not modelled in the simulation is the variation in the amount of scintillation light which reaches the photodetector given the longitudinal position in the crystal at which the light is emitted. Typically the peak position of the amount of scintillation light for electron showers is earlier in the crystal than photon showers and this is assumed as identical in the simulation. A single correlated nuisance parameter is applied to account for this effect.

Finally, one further systematic is applied to account for improvements made in more recent versions of the simulation which model electron bremsstrahlung at low  $p_T$  and photon conversion at high  $p_T$  much better. An additional correlated nuisance parameter is applied to account for this effect.

### Photon preselection efficiency

The efficiency of the preselection described in Sec. 4.5 is measured using  $Z \rightarrow e^+e^-$  decays in data and MC simulation. The MC efficiency is then corrected such that it

matches the data and this incurs a systematic uncertainty which gets applied as an overall normalisation on the expected signal event yield in each analysis category.

### Photon identification BDT

The photon identification BDT output (described in Sec. 5.1.2) is an important input to the diphoton BDT which is used to classify events. A shift in the photon ID BDT output, due to inaccurate simulation in the training samples, can have a direct effect on the output of the diphoton event BDT which in turn will lead to an event being misclassified; either it will fail the lower edge cut or it will go into a different analysis category. By examining the variation in the photon ID output for  $Z \rightarrow e^+e^-$  electrons,  $Z \rightarrow \mu^+\mu^- \gamma$  photons and diphoton events in data with  $m_{\gamma\gamma} > 160$  GeV it is found that applying a systematic shift of 0.01 to the output covers any discrepancies between data and MC simulation. This shift is applied and then propagated through to the diphoton BDT which in turn results in an uncertainty implemented as a relative yield change across event class (in other words: a category-migration).

### Photon energy resolution estimate

The most important input to the diphoton BDT, i.e. the variable with the most discriminating power, is the photon energy resolution estimate. As for the photon ID systematic above, the differences between data and MC are studied for  $Z \rightarrow e^+e^-$  decays,  $Z \rightarrow \mu^+\mu^- \gamma$  decays and high  $p_T$  photons. A systematic variation of the nominal value is applied and propagated through to the diphoton BDT output, after which the uncertainty is implemented as an analysis category-migration.

## 6.3.3 Systematic uncertainties related to diphoton events

### Vertex efficiency

The vertex efficiency in MC events is corrected by a scale factor obtained from the vertex efficiency ratio between data and MC as measured in  $Z \rightarrow \mu^+\mu^-$  events where the muon tracks are ignored (to replicate the situation with photons). The uncertainty associated with this is implemented by changing the relative fraction of the right/wrong vertex signal shape.



### Trigger efficiency

The trigger efficiency in MC events is corrected to match the data using  $Z \rightarrow e^+e^-$  events with a tag and probe technique[57]. The uncertainty associated to this is implemented as an overall normalisation change in each category.

### Mass scale non-linearity

Another important consideration when using the  $Z \rightarrow e^+e^-$  decay as a calibration and control tool in the analysis is not only the differences between electrons and photons but also the difference in the mass scale between the  $Z$  ( $m_Z = 91.2$  GeV[4]) and a Higgs boson at  $m_H = 125$  GeV. Although the absolute difference is not important, the relative differences between data and MC simulation are. This effect is measured by comparing the energy measured from the electromagnetic shower alone and the  $p_T$  from the tracker of electrons as a function of the scalar  $E_T$  sum of the two electrons in  $Z \rightarrow e^+e^-$  decays. This uncertainty is applied as a global mass shift, which has a correlated effect on the signal position in each category.

### Uncertainty on $Z$ mass

Given that most of the systematics on the energy scale are obtained from studies on the  $Z$  decay to electrons, the uncertainty on the  $Z$  mass measured in this decay is also included as a systematic. It has a correlated effect across all categories.

### Luminosity measurement

The luminosity measurement and its uncertainty is a CMS universal value and is described in Refs.[46, 47].

### Theoretical cross section and branching ratio

The uncertainties on the theoretically predicted SM Higgs cross section and its branching fraction to two photons are implemented following the recommendations of the LHC Higgs Cross Section Working Group[25].

### 6.3.4 Systematic uncertainties related to production mode tagged classes

#### Jet tagging efficiency

One of the largest uncertainties for measurement of the respective couplings to fermions and bosons of the Higgs boson originates from the complexity in isolating events produced by gluon fusion and those produced by vector boson fusion. As described in Sec. 5.2.1 this is done by tagging jets characteristic of  $VBF$  Higgs decays. However there is a large uncertainty involved when additional jets are produced by gluon fusion-induced Higgs decays. Using the Stewart-Tackmann procedure[76] the uncertainty on the yield of gluon fusion events in the  $VBF$  tagged event classes is calculated and implemented as a category-migration systematic. There are further effects from the jet energy scale, jet identification efficiency and the efficiency of rejecting jets incorrectly identified as pileup jets which are taken from Refs.[59, 77]. These affect the relative yields of the signal expectation between the  $VBF$  categories and also from the  $VBF$  categories as a whole to the inclusive categories.

#### Lepton, $\cancel{E}_T$ and $b$ -tagging efficiency

Additional normalisation uncertainties are applied to the exclusive mode categories which account for the efficiency of reconstructing and tagging leptons,  $\cancel{E}_T$  and  $b$ -jets.

#### $t\bar{t}H$ multijet rate

The effect of the jet identification efficiency on the  $t\bar{t}H$  multijet category. Although this systematic is quite large it only has an effect on the expected yield in this one category.

### 6.3.5 Summary

A summary of all the systematic uncertainties is shown in Table 6.2. A summary of the systematic uncertainties which affect the signal strength measurement, and the size of their effect, is shown in Table 6.3. A summary of the systematic uncertainties which affect the mass measurement, and the size of their effect, is shown in Table 6.4.

## 6.4 Statistical interpretation of the data

To aid with statistical interpretation of the data it is useful to define two hypotheses with which the data can be compared. These are the background only model and the signal plus background model, known as the null hypothesis,  $H_0$ , and the alternate hypothesis,  $H_1$ , respectively. The background model is obtained from the data by profiling the different parametrisation choices, as per the “envelope” method described in Sec. 6.2.1. The benchmark used for the signal model is the SM Higgs expectation as described in Sec. 6.1 which can be expressed in terms of two physical parameters of interest; the signal strength relative to the SM expectation,  $\mu = \sigma/\sigma_{\text{SM}}$ , and the signal position or Higgs mass,  $m_H$ . The likelihood function for statistical interpretation can be written as,

$$\mathcal{L}(\mu, m_H; \theta, \theta_b, \theta_d | m_{\gamma\gamma}) = \mu \cdot \mathbf{S}(m_{\gamma\gamma} | \mu, m_H; \theta) + \mathbf{B}(m_{\gamma\gamma} | \theta_b, \theta_d), \quad (6.15)$$

where  $m_{\gamma\gamma}$  is the diphoton invariant mass,  $\theta$  are continuous nuisance parameters affecting the signal model,  $\theta_b$  are continuous nuisance parameters affecting the background shape,  $\theta_d$  are discrete nuisance parameters dictating the background function and  $\mathbf{S}$  and  $\mathbf{B}$  are the probability distribution functions for the signal and background respectively. In the case of the SMVA there is no dependence on  $m_{\gamma\gamma}$  as events are simply counted in the analysis bins. It can be seen that  $\mu$  is a continuous parameter which represents the size of the fitted signal. The null hypothesis, or background only model, is a simple case of the full model where  $\mu = 0$ .

When determining the best fit values of parameters in our model and their errors, the full likelihood function, as expressed in Eq. 6.15, is fit to the data and the  $-2\Delta\text{LL}$  is scanned as a function of the parameter(s) of interest. This is a standard statistical procedure for measuring the value and error of a model parameter[69]. An important measurement to make of the Higgs boson is its signal strength, relative to the SM expectation, when considering fermionic production modes ( $ggH$  and  $t\bar{t}H$ ) and bosonic production modes ( $VBF$  and  $VH$ ) separately. This can help to ascertain the relative coupling strength of the Higgs to fermions and bosons. Given that the signal model is already expressed as a sum over the different production processes one can redefine the likelihood in this case so that the signal strength is split into two production dependent

parameters,  $\mu_{ggH,ttH}$  and  $\mu_{VBF,VH}$ :

$$\begin{aligned} \mu \cdot \mathbf{S}(m_{\gamma\gamma}|\mu, m_H; \theta) = & \mu_{ggH,ttH} \cdot \left( \mathbf{S}_{ggH}(m_{\gamma\gamma}|\mu, m_H; \theta) + \mathbf{S}_{ttH}(m_{\gamma\gamma}|\mu, m_H; \theta) \right) \\ & + \mu_{VBF,VH} \cdot \left( \mathbf{S}_{qqH}(m_{\gamma\gamma}|\mu, m_H; \theta) + \mathbf{S}_{VH}(m_{\gamma\gamma}|\mu, m_H; \theta) \right) \end{aligned} \quad (6.16)$$

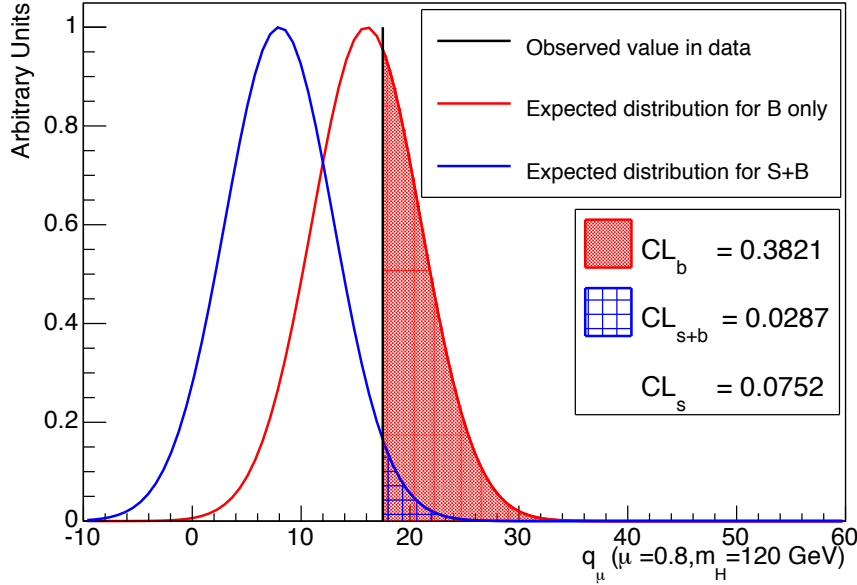
where  $\mathbf{S}_{\text{proc}}$  represents the signal model for the process “proc” only.

There are two important statistical tests that are performed aside from ascertaining the best fit values of parameters in the model. The first is an exclusion test designed to reject the alternate hypothesis in data. The second is a probability, or  $p$ -value, test designed to ascertain the likelihood of the null hypothesis fluctuating to give a signal. This requires the definition of a test statistic,

$$q_\mu = \begin{cases} -2 \ln \frac{\mathcal{L}(\text{data}|\mu, m_H, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, m_H, \hat{\theta}_{\hat{\mu}})} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} < 0 \end{cases}, \quad (6.17)$$

where  $\hat{\theta}_\mu$  and  $\hat{\theta}_{\hat{\mu}}$  represent the nuisance parameters at their best fit values given a particular value of  $\mu$  and given  $\mu$  is freely floating and at its global best fit value  $\hat{\mu}$ [78]. The statistical tests which require  $q_\mu$  are carried out at a specific hypothesised Higgs mass and so  $m_H$  is set to a particular value when defining the test statistic. It can be shown that  $q_\mu$  is the most powerful test of hypothesis  $H_0$  against hypothesis  $H_1$  and this is known as the *Neymann-Pearson Lemma*[69].

For exclusion limits the  $CL_s$  method is used, which is designed to give less stringent limits in situations where there is little discriminatory power between the null and alternate hypotheses[79]. The  $CL_s$  exclusion requires calculation of two  $p$ -values given the test statistic distribution of the two hypotheses is known, as shown in Eqs. 6.18- 6.20. A common way of obtaining this distribution is by generating pseudo-data under each hypothesis. However asymptotic approximations are used to avoid the computational overhead of generating toy experiments[78]. A one-sided upper limit is determined on  $\mu$  by enforcing the constraint that  $\hat{\mu} \leq \mu$  and the exclusion power of the limit is given by



**Figure 6.11:** A demonstration of the test statistic distribution for the background only model (red) and the signal plus background model (blue) and the observed value in data (black line). The exclusion power is given by  $1 - CL_s$ , in this case values of  $\mu > 0.8$  are excluded at 92.48% confidence.

$1 - CL_s$ . The definition of  $CL_s$  is

$$CL_s = \frac{CL_{s+b}}{CL_b}, \quad (6.18)$$

$$CL_{s+b} = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mu = \hat{\mu}) dq_\mu, \quad (6.19)$$

$$CL_b = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mu = 0) dq_\mu, \quad (6.20)$$

where  $q_\mu^{obs}$  is the observed value of the test statistic in data and  $f(q_\mu | \mu)$  is the distribution of the test statistic for the hypothesis with value  $\mu$ . The  $CL_s$  exclusion method is demonstrated by an example in Fig. 6.11. In this case a SM Higgs boson with values of signal strength  $\mu > 0.8$  is excluded at 92.48%. In reality the desired exclusion level is 95% and consequently the value of  $\mu$  is adjusted until the  $CL_s$  reaches 0.05 and then all values of  $\mu$  greater than or equal to this value are considered as excluded. When referring to the “expected” exclusion (as opposed to the “observed” exclusion just explained) the observation value is considered as being at the mean of the null hypothesis test statistic distribution.

When quantifying the significance of an observed excess the test statistic  $q_0$  is used which is obtained by setting  $\mu = 0$  in Eq. 6.17. The requirement that  $\hat{\mu} \geq 0$  ensures that only positive excesses are considered significant. The probability that the background-only hypothesis is rejected in favour of the signal plus background hypothesis is given in terms of the  $p$ -value,  $p_0$ , defined as

$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0|\mu = 0) dq_0. \quad (6.21)$$

Typically a value of  $p_0 \leq 1.3 \times 10^{-3} = 3\sigma$  is enough to claim observation of new physics while  $p_0 \leq 2.87 \times 10^{-7} = 5\sigma$  is enough to claim a discovery of new physics.

In the spin analysis the signal model is parametrised in terms of  $\mu$ ,  $m_H$  and two additional parameters,  $x$  and  $f_{q\bar{q}}$ , which dictate the amount of signal from spin-2 and the fraction of spin-2 from  $q\bar{q}$  production. The signal model parametrisation can be written as

$$\mathbf{S}(m_H, f_{q\bar{q}}, x; \theta) = (1 - x) \cdot \mathbf{S}_{SM}(m_H; \theta) + x \cdot \left[ f_{q\bar{q}} \cdot \mathbf{S}_{q\bar{q}}(m_H; \theta) + (1 - f_{q\bar{q}}) \mathbf{S}_{gg}(m_H; \theta) \right]. \quad (6.22)$$

**Table 6.2:** A summary of all the systematic uncertainties considered in the analysis.

Uncertainty Name			Uncertainty Value (%)			
Per photon uncertainties						
			Barrel		Endcap	
			low $\eta$	high $\eta$	low $\eta$	high $\eta$
Energy scale	7 TeV	$R_9 \geq 0.94$	0.05	0.10	0.10	0.05
		$R_9 < 0.94$	0.05	0.05	0.10	0.10
	8 TeV	$R_9 \geq 0.94$	0.05	0.10	0.10	0.05
		$R_9 < 0.94$	0.05	0.05	0.10	0.10
Energy resolution (constant)	7 TeV	$R_9 \geq 0.94$	0.08	0.21	0.23	0.09
		$R_9 < 0.94$	0.07	0.11	0.31	0.13
	8 TeV	$R_9 \geq 0.94$	0.06	0.09	0.10	0.08
		$R_9 < 0.94$	0.05	0.10	0.12	0.06
Energy resolution (stochastic)	8 TeV	$R_9 \geq 0.94$	0.16	0.22	-	-
		$R_9 < 0.94$	0.16	0.07	-	-
Mismodelling of tracking material			0.04-0.34		0.22-0.34	
Non-uniformity in light emission peak					0.05	
Mismodelling of shower shape (GEANT4)					0.05	
Photon preselection efficiency			1.0		2.6	
Photon identification BDT					0.5	
Photon resolution estimate					10.0	
Per event uncertainties						
Vertex efficiency					1.5	
Trigger efficiency					1.0	
Non-linearity in mass scale					0.1-0.2	
Uncertainty on $Z$ mass					0.01	
Luminosity measurement	7 TeV				2.2	
	8 TeV				2.6	
Theory (cross section)	7 TeV				2.1-12.3	
	8 TeV				2.6-12.3	
Theory (branching ratio)					5.0	
Uncertainties related to tagged classes						
Jet energy scale						
<i>up to 10% event migration between dijet classes and untagged classes</i>						
Theory uncertainties on dijet production from gluon fusion						
<i>up to 30% event migration between dijet classes and untagged classes</i>						
Muon efficiency					0.2-0.6	
Electron efficiency					0.2-0.4	
$\cancel{E}_T$ efficiency					1-4	
$b$ -tagging efficiency					1-2	
$t\bar{t}H$ multijet rate					13-30	

**Table 6.3:** Magnitude of the uncertainty on the signal strength,  $\sigma/\sigma_{\text{SM}}$ , induced by the systematic uncertainties on the signal.

Systematic uncertainty	Uncertainty (%)
	on $\sigma/\sigma_{\text{SM}}$
Energy scale and resolution corrections	0.02
Uncertainty from resolution estimate and photon identification BDT	0.06
Other experimental uncertainties	0.04
Theoretical uncertainties	0.11
All systematic uncertainties on the signal model	0.13

**Table 6.4:** Magnitude of the uncertainty on the signal position,  $m_H$ , induced by the systematic uncertainties on the signal.

Systematic uncertainty	Uncertainty (GeV)
	on $m_H$
Energy scale and resolution corrections	0.05
Non-linearity extrapolation from $Z$ -boson scale to Higgs scale	0.10
Differences between electrons and photons	0.11
All systematic uncertainties on the signal model	0.16



# Chapter 7

## Results

In the MFM analysis the data in all 25 analysis categories (11 at 7 TeV, 14 at 8 TeV) is simultaneously fitted with a background and signal component. The background normalisation and shape parameters (which includes the different function choices) are completely free floating, i.e. there is no constraint on their values. The signal shape is allowed to vary according to the systematic uncertainties described in Sec. 6.3. The parametrisation of the signal yield is such that, although the overall rate is floating, the fraction in each category is required to match the prediction of a SM Higgs boson (modulo the nuisance parameters). For some of the coupling measurements these signal constraints are relaxed such that the signal strength through bosonic production modes ( $VBF, VH$ ) and through fermionic production modes ( $ggH, t\bar{t}H$ ) are allowed to float independently.

Where appropriate the results of the MFM and SMVA are shown next to each other. As the SMVA is a cross-check to the background model and categorisation scheme, only results on exclusion power, observation significance and signal strength measurement are provided for the SMVA. The results of couplings measurements and compatibility across channels and production mechanisms are shown only for the MFM.

The first section shows the results of the best fit signal plus background model to data. In the second section the results of the statistical hypothesis tests are shown: the exclusion limit of a SM Higgs boson and the local probability of a signal-like background fluctuation ( $p$ -value test). These establish an independent discovery of a Higgs-like resonance near 125 GeV. The last two sections focus on understanding some of the properties of the observed state with respect to the SM expectation: measurements of

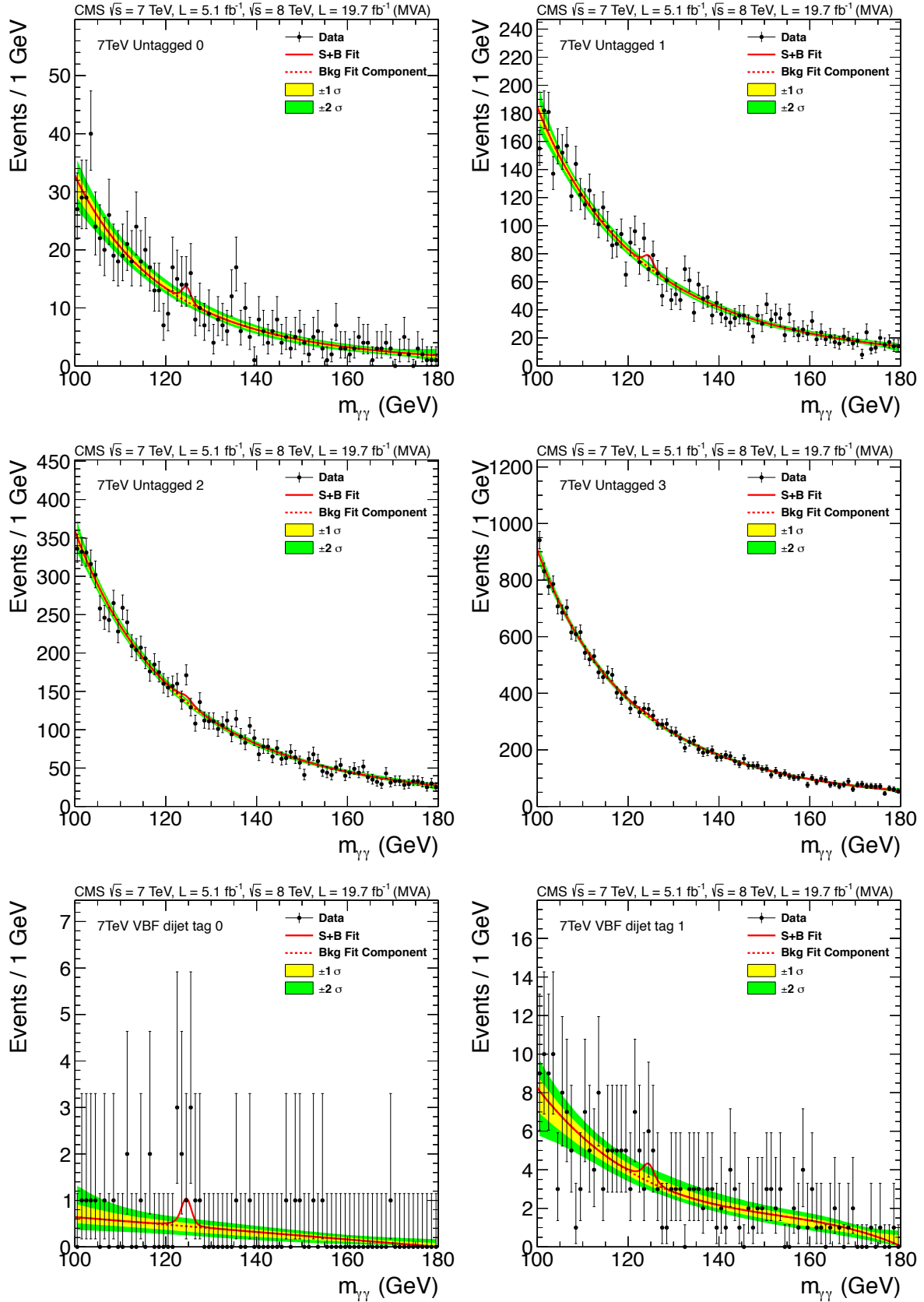
its mass and coupling parameters and hypothesis tests comparing various spin models to the data.

## 7.1 Best fit model to data

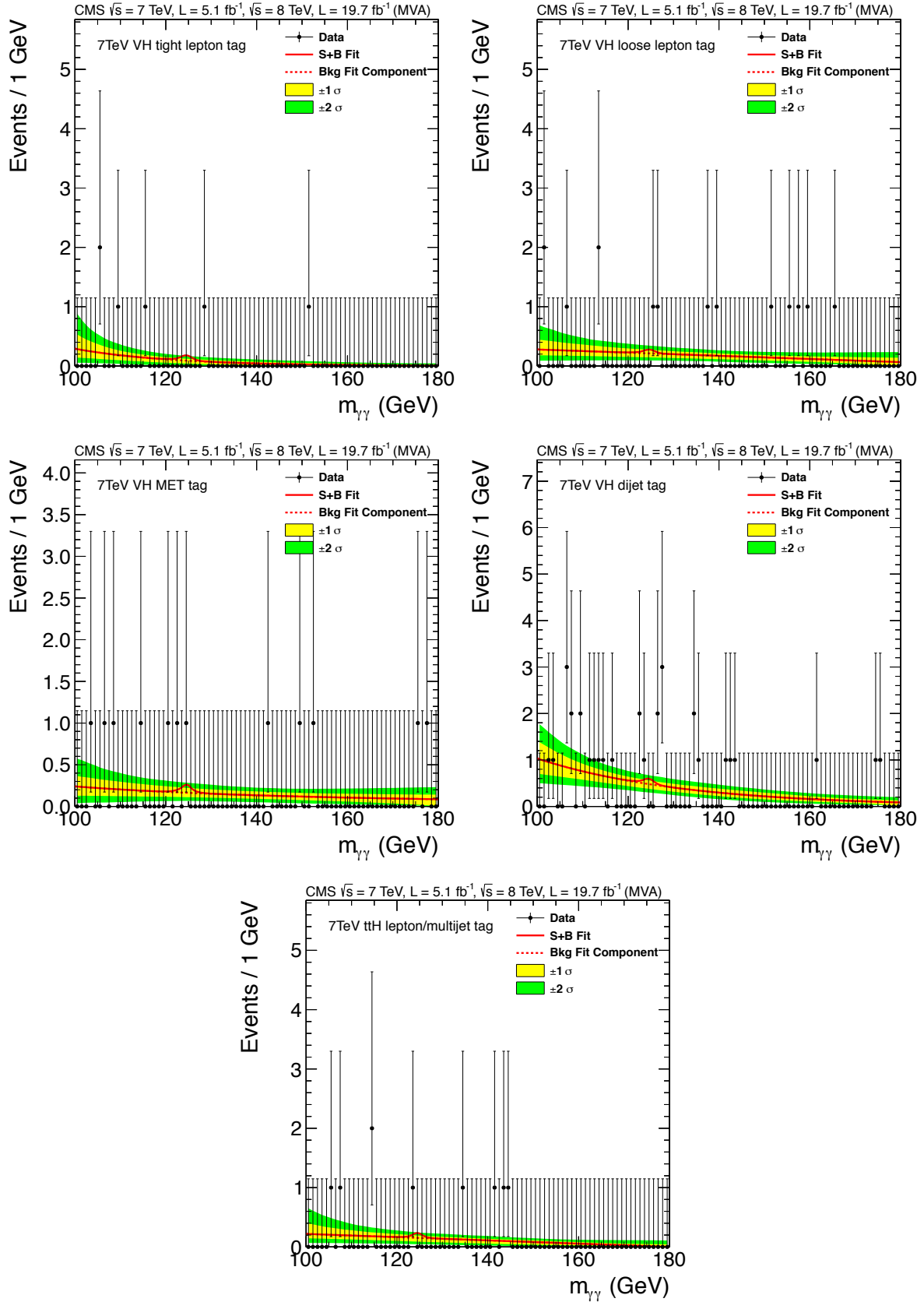
This section shows the best fit result of the signal plus background model when floating both the signal strength,  $\mu = \sigma/\sigma_{\text{SM}}$ , and the signal position,  $m_H$ , whilst fitting simultaneously across all of the analysis categories using the nominal MFM analysis. Figures 7.1-7.5 show the best fit result in each analysis category. Figures 7.6 and 7.7 show the fully inclusive distribution with the full signal plus background model and a data-background residual underneath. It is apparent from these two plots that a substantial signal is observed near  $m_H = 125$  GeV.

## 7.2 Exclusion limits and $p$ -value

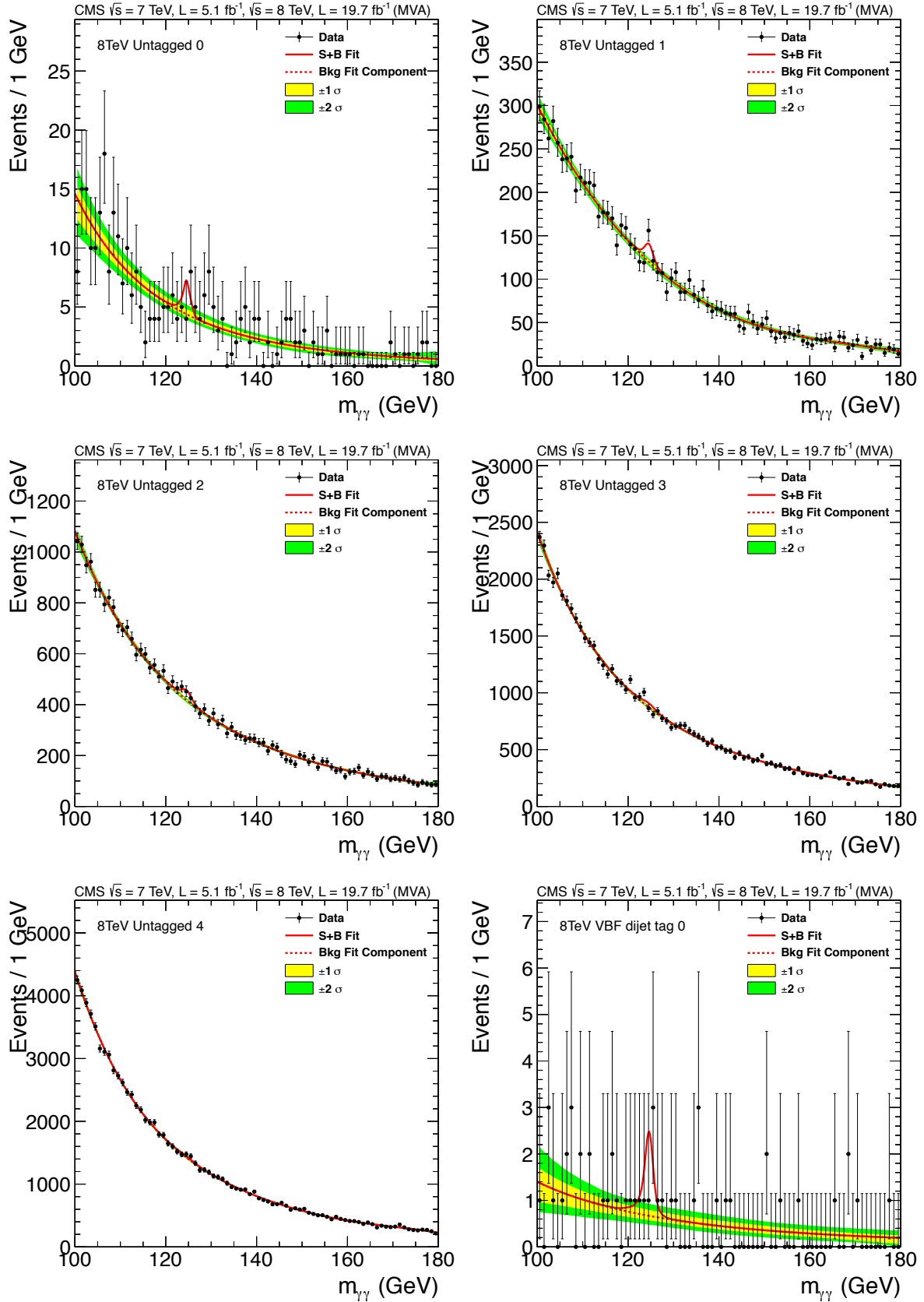
The exclusion limit for a SM Higgs boson is shown in Fig. 7.8 for the MFM analysis (left) and the SMVA analysis (right) for the combined 7 and 8 TeV dataset. In this figure, the test statistic used is the one defined in Eq. 6.17. The signal position,  $m_H$ , is set to a given value and the signal strength,  $\mu$ , is altered until the required value of  $CL_s$ , in this case 5%, is obtained. This allows determination of the signal strength excluded at 95% C.L. for a given signal mass. This plot shows the cross section, relative to the SM prediction, excluded at 95% confidence using the  $CL_s$  method as a function of the Higgs mass,  $m_H$ , in steps of 1 GeV. This has become a standard plot for Higgs searches and aims to determine what cross sections can be ruled out for a Higgs boson at a given mass. The black dashed line, with the green and yellow bands, represents the expected exclusion, and the associated  $\pm 1\sigma$  and  $\pm 2\sigma$  error, if exactly the mean of the null hypothesis is observed. The solid black line represents the actual observation in data. For values of  $m_H$  where the solid black line falls below the solid red line at 1, the SM Higgs boson is excluded at a confidence level of 95% or higher. It is clear that a SM Higgs boson is excluded at all values of  $m_H$  apart from near 125 GeV where there is a large excess (well above the  $2\sigma$  band). It can be seen that both the nominal MFM analysis and the cross-check SMVA have similar exclusion power (shown by the position of the dashed line and coloured bands) and a similar observation.



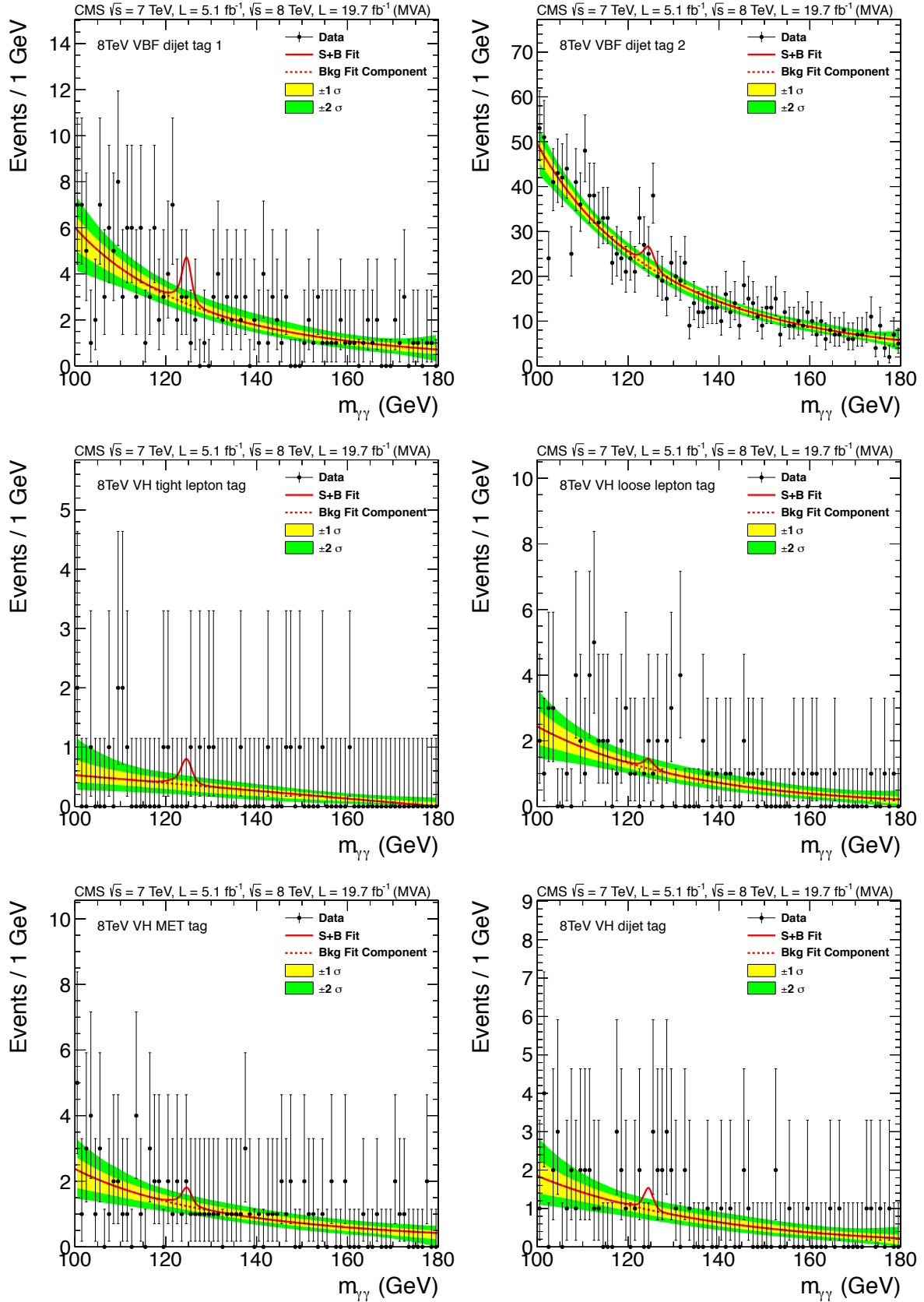
**Figure 7.1:** The diphoton invariant mass distribution in data (black points) with the best fit signal plus background overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). Shown for the untagged and dijet categories in the 7 TeV dataset.



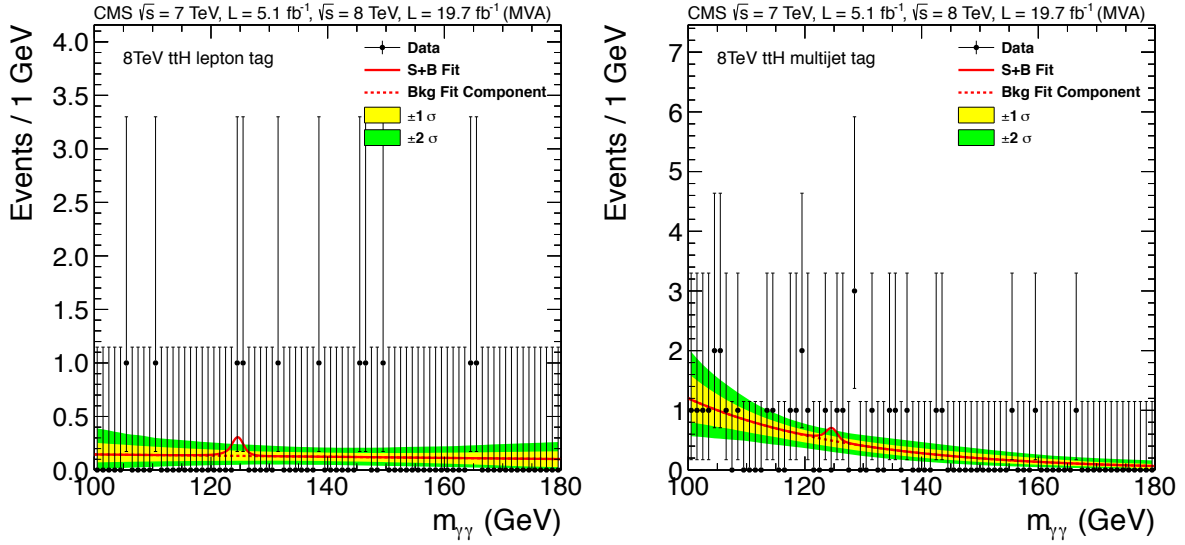
**Figure 7.2:** The diphoton invariant mass distribution in data (black points) with the best fit signal plus background overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). Shown for the  $VH$  and  $t\bar{t}H$ -tagged categories in the 7 TeV dataset.



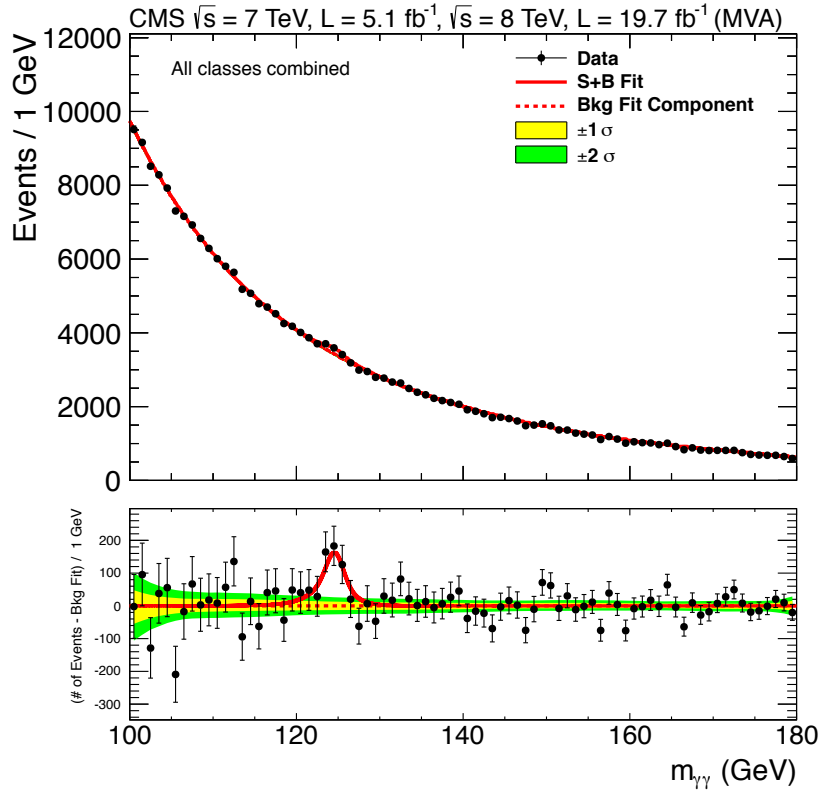
**Figure 7.3:** The diphoton invariant mass distribution in data (black points) with the best fit signal plus background overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). Shown for the untagged and dijet 0 categories in the 8 TeV dataset.



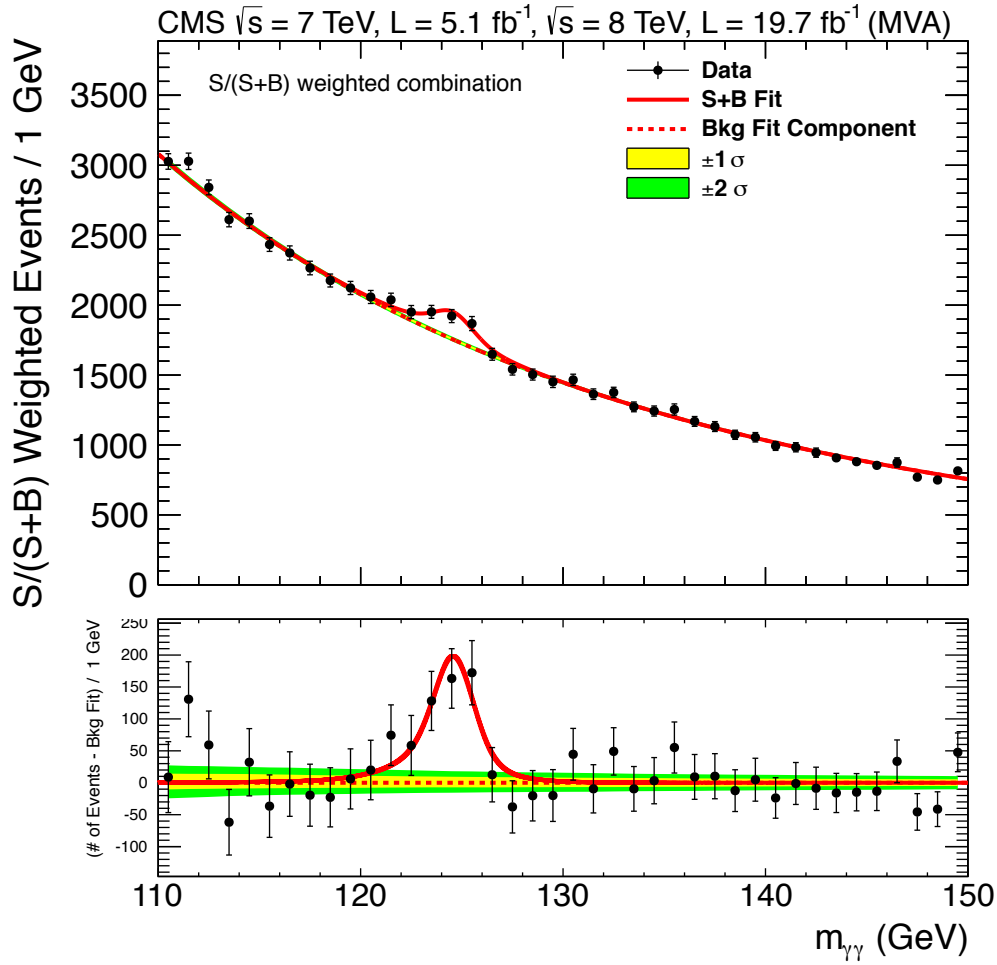
**Figure 7.4:** The diphoton invariant mass distribution in data (black points) with the best fit signal plus background overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). Shown for the dijet 1,2 and  $VH$ -tagged categories in the 8 TeV dataset.



**Figure 7.5:** The diphoton invariant mass distribution in data (black points) with the best fit signal plus background overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). Shown for the  $t\bar{t}H$ -tagged categories in the 8 TeV dataset.

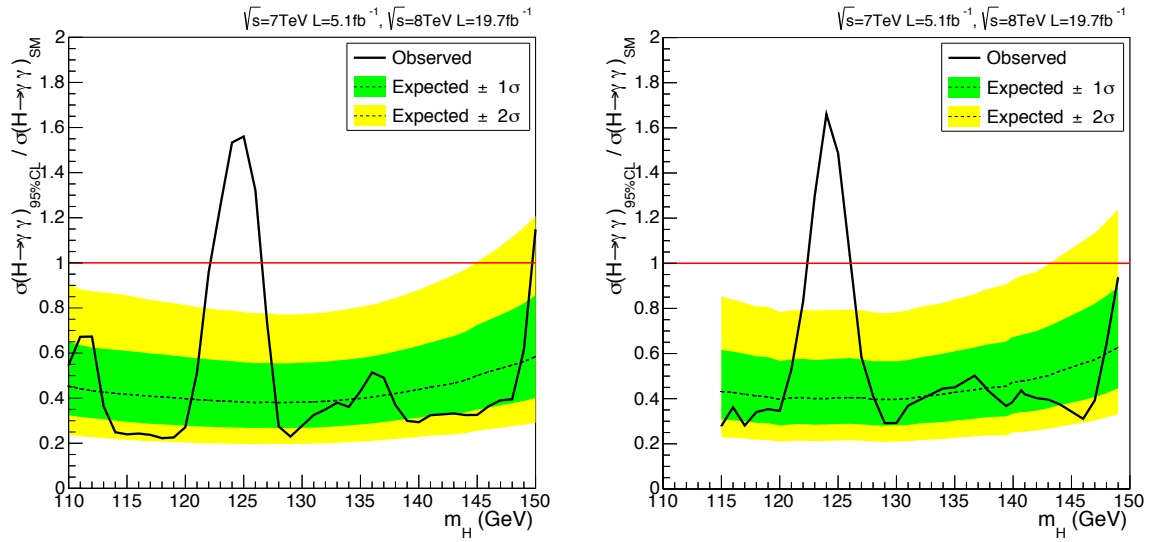


**Figure 7.6:** The diphoton invariant mass distribution for all analysis events in the range  $100 < m_{\gamma\gamma} < 180$  GeV in data (black points) with the combined best fit signal plus background model overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). The data minus background residual is shown at the bottom.

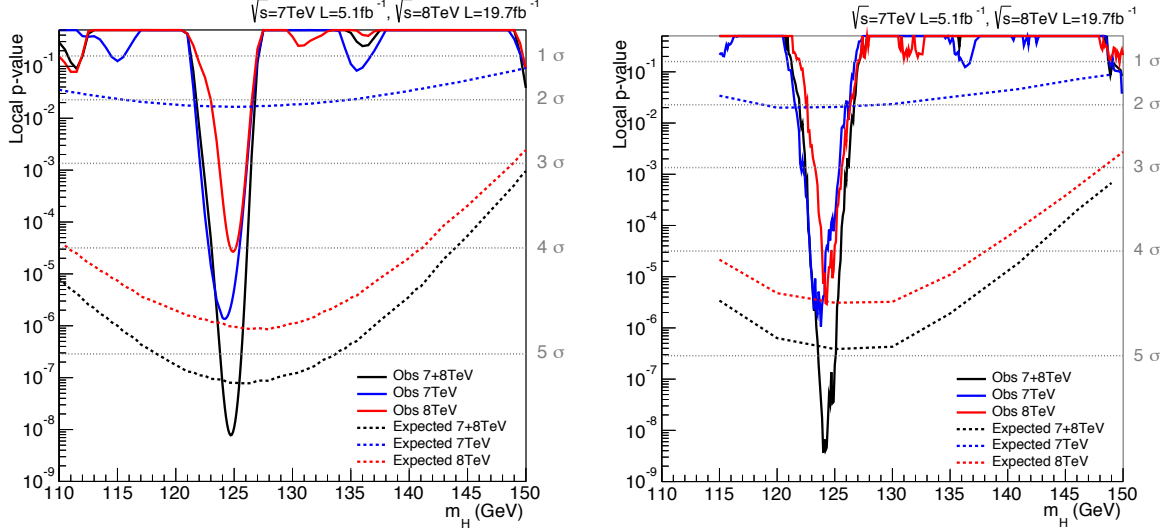


**Figure 7.7:** The diphoton invariant mass distribution for all analysis events in the analysis search range  $110 < m_{\gamma\gamma} < 150$  GeV in data (black points) with the combined best fit signal plus background model overlaid (solid red line), including the background only component (dashed red line) and the uncertainty on the background shape (yellow and green bands). In this plot the data, signal and background models in each analysis category are given a weight according to the expected  $S/(S+B)$  ratio in  $\pm\sigma_{\text{eff}}$  of the best fit mass. The plot is then normalised such that the integral under the signal peak gives the same number of events as the non-weighted equivalent. This plot helps to show the fitted signal by upweighting categories with higher sensitivity. The data minus background residual is shown at the bottom.



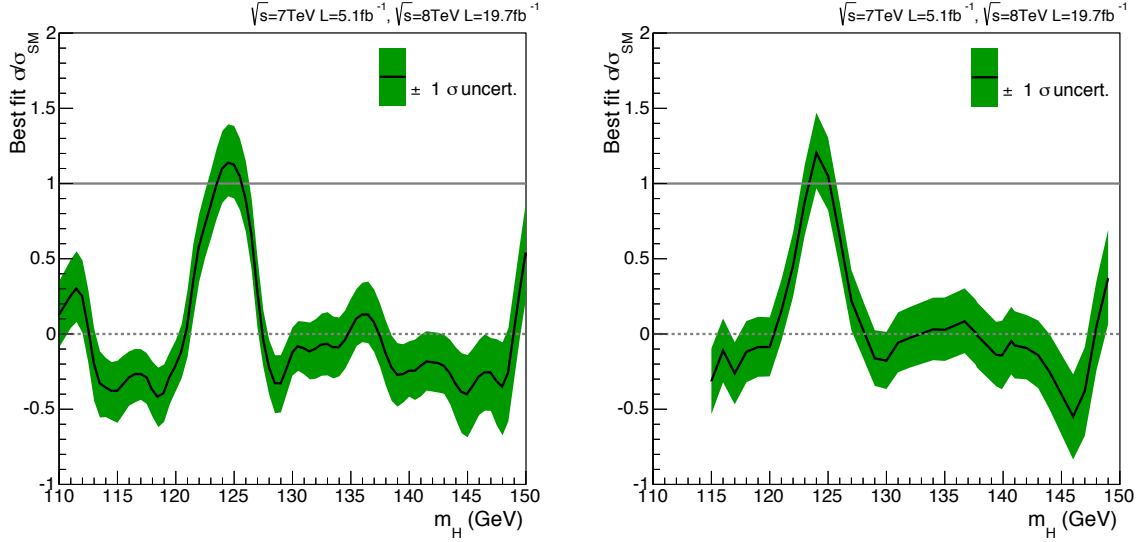


**Figure 7.8:** The expected and observed exclusion limits for a SM Higgs boson at 95% confidence level. The dashed line shows the expected exclusion if exactly the mean of the null hypothesis is observed with its error at  $1\sigma$  (green) and  $2\sigma$  (yellow). The solid black line shows the observed exclusion. The results are shown for the nominal MFM analysis (left) and the cross-check SMVA (right) when combining both the 7 and 8 TeV datasets. A SM Higgs boson is disfavoured everywhere, at 95% confidence level or higher, apart from where there is a large excess around 125 GeV.



**Figure 7.9:** The expected (dashed lines) and observed (solid lines) local  $p$ -value to reject the background only hypothesis as a function of the hypothesised Higgs mass,  $m_H$ . The 7 TeV (blue lines) and 8 TeV (red lines) results are shown separately along with the combination (black lines). The results are shown for the nominal MFM analysis (left) and the cross-check SMVA (right). The observed  $p$ -value in the MFM at the most significant point is  $5.7\sigma$  ( $m_H = 124.7$  GeV) given a SM expectation of  $5.2\sigma$ .

In order to quantify the significance of this excess the test statistic defined in Eq. 6.21 is used. This is employed to determine the probability, the local  $p$ -value, that the background can fluctuate to give an observed signal by comparing the likelihoods for the null and alternate hypothesis at a given Higgs mass. The observed local  $p$ -value is shown as a function of the hypothesised Higgs mass,  $m_H$ , in Fig. 7.9 for both the 7 TeV (blue line) and 8 TeV (red line) datasets separately and the combination (black line). It is clear there is a significant excess in both subsets of data at approximately the same mass and the quantity of the excess is apparent in both the MFM and SMVA analyses. The local  $p$ -value for the MFM using the combined 7 and 8 TeV datasets at  $m_H = 124.7$  GeV (the mass with the most significant excess) equates to a significance of  $5.7\sigma$  where  $5.2\sigma$  is the expectation for a SM Higgs boson. This constitutes a standalone discovery of a Higgs-like resonance around 125 GeV.

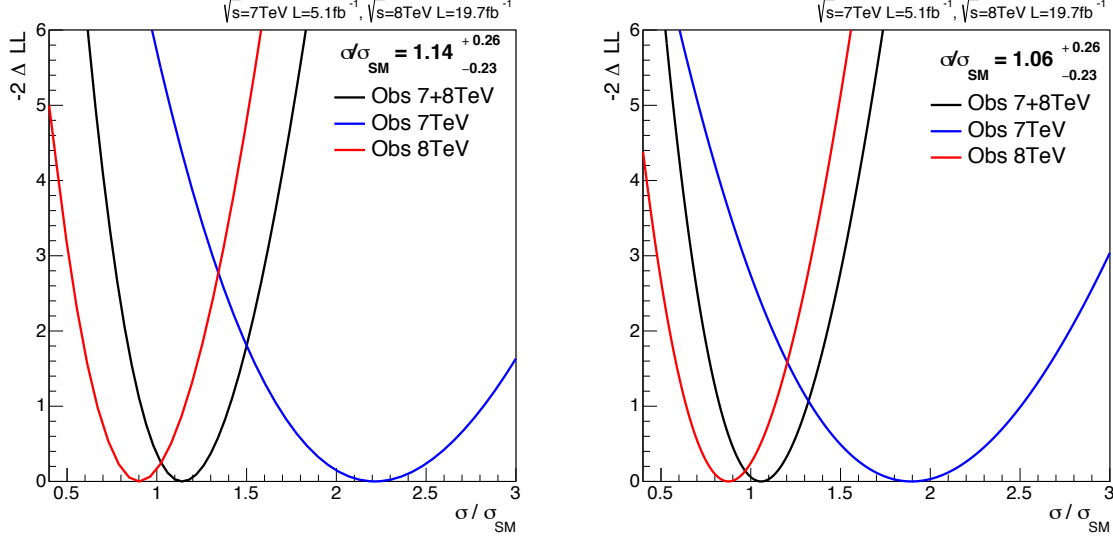


**Figure 7.10:** The observed best fit value of the signal strength modifier,  $\mu = \sigma/\sigma_{\text{SM}}$ , as a function of the hypothesised Higgs mass  $m_H$  when using the combined 7 and 8 TeV datasets. The  $\pm 1\sigma$  error band is shown as the green band. Results are shown for the nominal MFM analysis (left) and the cross-check SMVA analysis (right).

### 7.3 Measurement of physical parameters

The best fit value of the signal strength modifier,  $\mu = \sigma/\sigma_{\text{SM}}$ , is shown as a function of the hypothesis Higgs mass,  $m_H$ , in Fig. 7.10 for the combined 7 and 8 TeV dataset. As expected this follows very closely the shape of the observed exclusion in Fig. 7.8 and shows that the observed boson is very compatible with a SM Higgs, i.e. the observed value of  $\mu$  is within  $1\sigma$  of the SM expectation,  $\mu = 1$ .

Figure 7.11 shows the one dimensional  $-2\Delta\text{LL}$  scan of  $\mu$  when the Higgs mass  $m_H$  is profiled. The hypothesis mass,  $m_H$ , is left floating in the fit as there is no *a priori* knowledge of its value. One can see that the results are very consistent between the MFM (left) and the SMVA (right). There is some distance between the measurement made using the 7 TeV dataset (blue line) and the 8 TeV dataset (red line) but they are consistent at  $< 2\sigma$  level and it is quite reasonable to imagine an upward fluctuation in the data at 7 TeV and a downward fluctuation at 8 TeV. When combining the two datasets the best fit value comes out very close to the SM expectation of  $\mu = 1$  so certainly an interesting area for future measurements with more data in the  $H \rightarrow \gamma\gamma$  channel is where this value goes. The best fit values of  $\sigma/\sigma_{\text{SM}}$  with their errors (and the mass at which the best fit is found)



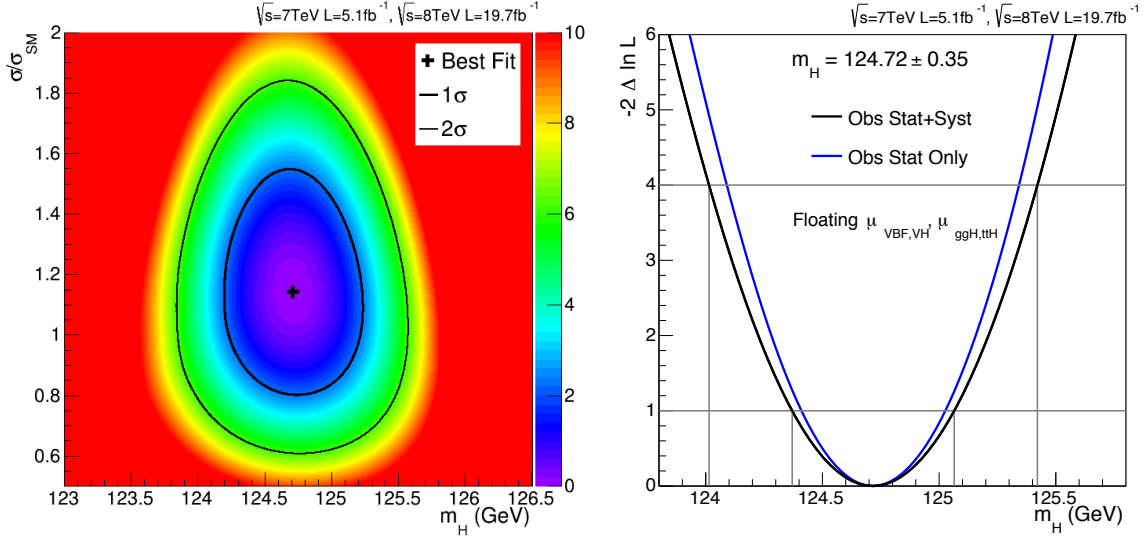
**Figure 7.11:** The one dimensional  $-2\Delta LL$  scan of the signal strength modifier,  $\mu = \sigma/\sigma_{\text{SM}}$ , when the hypothesis Higgs mass,  $m_H$ , is profiled in the fit. The results are shown for the 7 TeV (blue), the 8 TeV (red) and combined (black) datasets using the MFM analysis (left) and the SMVA analysis (right). The observed best fit value for the nominal MFM analysis is  $\sigma/\sigma_{\text{SM}} = 1.14^{+0.26}_{-0.23}$ .

**Table 7.1:** The values of the best fit signal strength,  $\sigma/\sigma_{\text{SM}}$ , when  $m_H$  is profiled, for the 7 TeV, 8 TeV, and combined datasets. The value of  $m_H$  at which the best fit occurs is also given.

	$\sigma/\sigma_{\text{SM}}$	$m_H$ (GeV)
7 TeV	$2.22^{+0.60}_{-0.54}$	124.2
8 TeV	$0.90^{+0.25}_{-0.23}$	124.9
7 + 8 TeV	$1.14^{+0.26}_{-0.23}$	124.7

are summarised for each of the datasets using the MFM analysis in Table 7.1. The overall best fit of the signal strength is  $\sigma/\sigma_{\text{SM}} = 1.14^{+0.26}_{-0.23} \left[ {}^{+0.21}_{-0.21}(\text{stat}) {}^{+0.13}_{-0.09}(\text{theory}) {}^{+0.09}_{-0.05}(\text{syst}) \right]$ .

The following measurements focus on the properties of the observed signal and are consequently only shown for the MFM analysis. The two most important physical parameters to measure in the signal are the overall rate,  $\mu$ , and the mass,  $m_H$ . We have seen the one dimensional  $-2\Delta LL$  scan of the signal strength,  $\mu$ , when  $m_H$  is profiled. The result of doing a two dimensional  $-2\Delta LL$  scan in both the parameters simultaneously is shown on the left hand side of Fig. 7.12 for the combined 7+8 TeV dataset. For a standalone measurement of the mass of the observed particle it is undesirable to constrain the signal resonance to have exactly SM couplings and production rates. Consequently,



**Figure 7.12:** The two dimensional  $-2\Delta\text{LL}$  scan of  $\mu$  and  $m_H$  when fitting to the combined 7+8 TeV dataset is shown on the left. The best fit value is shown by the black cross and the  $1\sigma$  and  $2\sigma$  error contours shown by the solid lines. The one dimensional  $-2\Delta\text{LL}$  scan of the best fit mass  $m_H$  is shown on the right, when profiling over the relative signal production from fermionic and bosonic modes,  $\mu_{ggH,ttH}$  and  $\mu_{VBF,VH}$ . The statistical-only component is shown as the blue line and the statistical plus systematic as the black line.

when measuring the mass, the overall signal rate is allowed to scale in terms of the production from fermion couplings,  $\mu_{ggH,ttH}$ , and from boson couplings,  $\mu_{VBF,VH}$  (see Eq. 6.16). The one dimensional  $-2\Delta\text{LL}$  scan of the observed mass is shown on the right hand side of Fig. 7.12 for the combined 7+8 TeV dataset when profiling the  $\mu_{VBF,VH}$  and  $\mu_{ggH,ttH}$  signal strengths. It should be noted that in principle the best fit mass when scanning the two dimensions of  $\mu$  and  $m_H$  (Fig. 7.12 left) does not necessarily have to be the same as the mass when scanning one dimension  $m_H$  while floating two other parameters,  $\mu_{VBF,VH}$  and  $\mu_{ggH,ttH}$  (Fig. 7.12 right). However, in practice they come out almost identical because, as is shown below, the coupling strength parameters  $\mu_{VBF,VH}$  and  $\mu_{ggH,ttH}$  come out very close to the SM expectation. The observed best fit mass of the boson is found to be  $m_H = 124.72 \pm 0.35(\text{stat}) \pm 0.16(\text{syst})$  GeV.

Aside from the mass and signal strength of the observed particle it is relevant to study its couplings and whether the relative fractions produced by the different production modes are also compatible with the SM prediction. This is addressed in Fig. 7.13. One-dimensional  $-2\Delta\text{LL}$  scans of the relative couplings to fermions,  $\mu_{ggH,ttH}$ , and bosons,  $\mu_{VBF,VH}$ , are shown in the top right and top left plots respectively. It is clear that the excess of signal in 7 TeV dataset is apparent in both the fermionic and bosonic production

modes, with a large excess in the tight  $VBF$  category at 7 TeV driving the very high value of  $\mu_{VBF,VH}$  in the 7 TeV dataset. The slight deficit, with respect to the SM, in the 8 TeV dataset comes mainly from the bosonic production modes. Figure 7.13 also shows the equivalent two dimensional  $-2\Delta\text{LL}$  scan of both of these parameters simultaneously. It can be seen that the observation of  $\mu_{ggH,t\bar{t}H} = 1.13^{+0.37}_{-0.31}$ ,  $\mu_{qqH,VH} = 1.16^{+0.63}_{-0.57}$  is very compatible with the SM expectation. Furthermore, the signal strength can be divided by production mode as also shown in Fig. 7.13. It can be seen that the signal strength for all four production mechanisms is consistent with the SM expectation at 1. The Higgs mass,  $m_H$ , is profiled in all of these scans but constrained to be the same across each production mode.

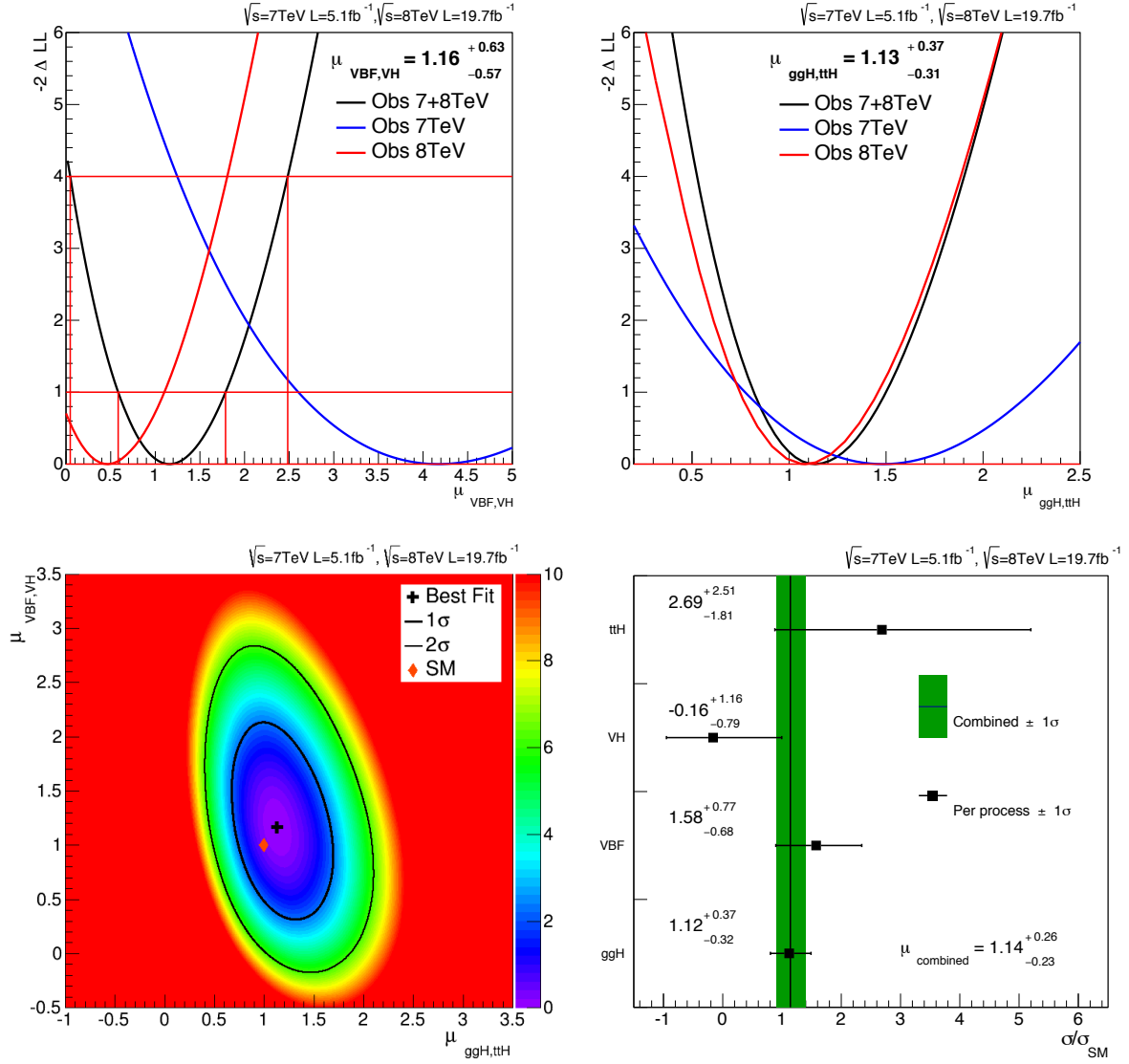
Figure 7.14 shows the breakdown of the extracted signal strength when separately fitted for each of the analysis categories and when grouping the categories by their topology into those which are untagged, those which are dijet (or  $VBF$ ) tagged,  $VH$  tagged or  $t\bar{t}H$  tagged. It is noticeable that the categories with the most sensitivity are the untagged categories, especially the “Untagged 1” and “Untagged 2” at 8 TeV, whilst the dijet categories also have considerable sensitivity because the signal to background ratio is high and the statistics are reasonable. The  $VH$  and  $t\bar{t}H$  tagged categories have the least sensitivity and whilst they do not contribute much to the error on the total signal yield they are important when measuring the couplings of the observed boson.

## 7.4 Spin

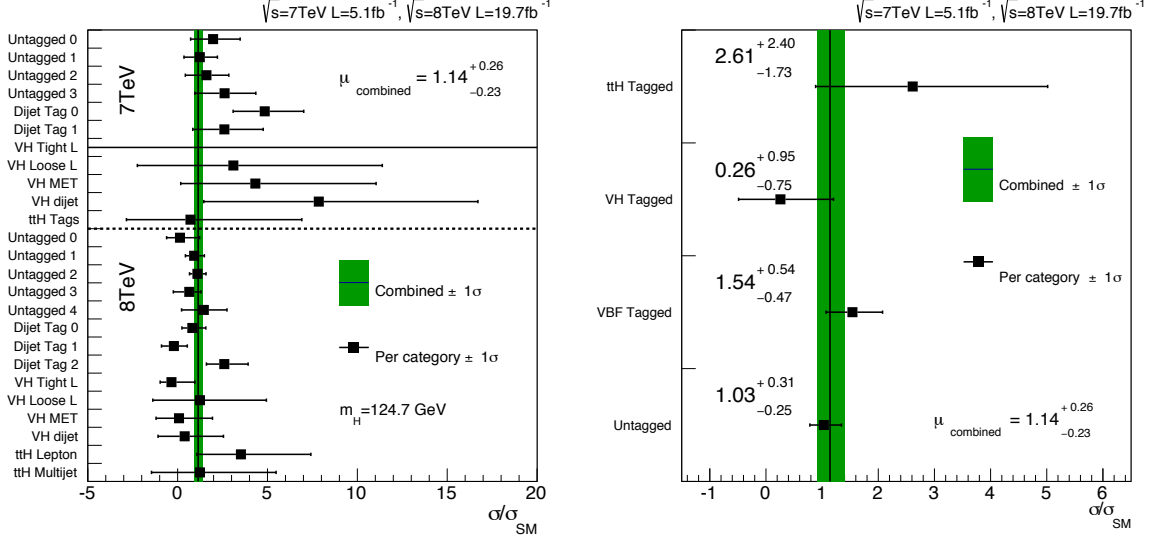
This section presents results of the spin analysis. The first part concentrates on extracting the differential signal strength, relative to the SM expectation, as a function of the decay angle,  $\cos(\theta_{\text{CS}}^*)$  (see Eq. 5.5). The second part describes the results of a statistical hypothesis test to exclude particular spin-2 models.

### 7.4.1 SM compatibility check

The signal yield,  $\mu = \sigma/\sigma_{SM}$ , is extracted independently in each of the  $|\cos(\theta_{\text{CS}}^*)|$  bins, simultaneously fitting over the  $\eta$  and  $R_9$  bins such that the relative yields in each of the  $\eta$  and  $R_9$  bins is constrained to that predicted by the SM. The result is shown in Figure 7.15 for the data and various  $2_m^+$  model expectations, where for the expectations a single representative toy is used, obtained using asymptotic formulae from Ref[78], and



**Figure 7.13:** The top row shows the one dimensional  $-2\Delta LL$  scan of the SM relative signal strength from bosonic production modes,  $\mu_{VBF,VH}$  (top left), and from fermionic production modes,  $\mu_{ggH,ttH}$  (top right) for the 7 TeV (red), 8 TeV (blue) and combined (black) datasets. The bottom left plot is the 2D  $-2\Delta LL$  scan of  $\mu_{VBF,VH}$  ( $y$ -axis) vs.  $\mu_{ggH,ttH}$  ( $x$ -axis) for the combined dataset observation. The best fit point is shown by the black cross whilst the  $1\sigma$  and  $2\sigma$  error intervals are shown by the solid lines. The SM expectation is the red diamond. It can be seen that the observation is very compatible with the SM prediction. The bottom right plot shows the value of the observed signal strength (black points) when floating the separate contributions from each production process with a common mass compared to the combined best fit value (green band). It can be seen that each individual coupling component is compatible with the observation and the SM expectation. The  $\chi^2$   $p$ -value for this compatibility is  $p(\chi^2) = 49\%$ .



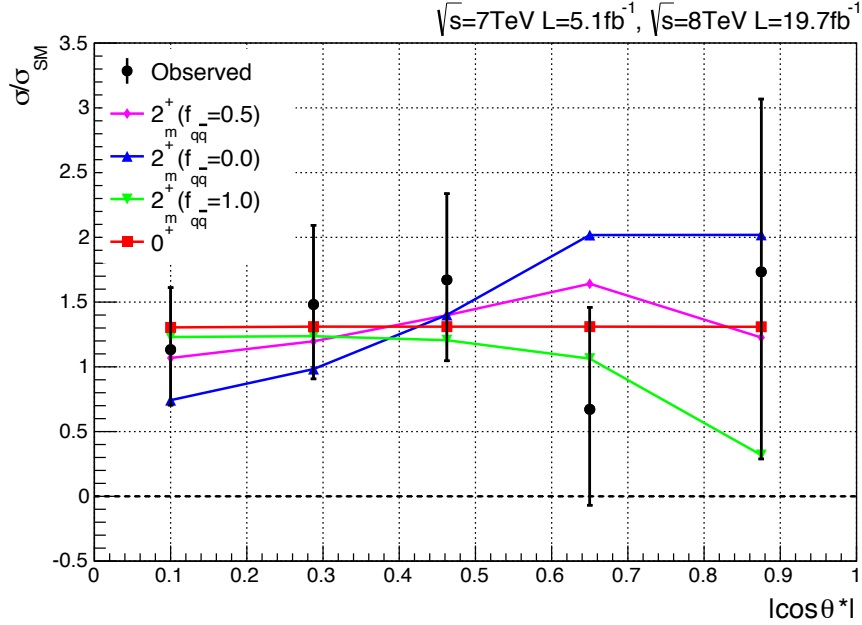
**Figure 7.14:** The left plot shows the signal strength breakdown when fitting the signal strength,  $\mu$ , in each analysis category separately. The compatibility is found to be  $p(\chi^2) = 74\%$ . The right plot shows the signal strength breakdown when fitting categories split by their topology (tag type). The compatibility is found to be  $p(\chi^2) = 49\%$ .

the normalisation is extracted from a fit to data. The concept of Fig. 7.15 is to compare the data distribution of  $|\cos(\theta_{cs}^*)|$  to the expectation of various different spin hypotheses. One would expect that the coloured lines on this figure replicate the generator level distributions shown in Fig. 5.10. This is the case for the SM  $0^+$  expectation, however the final point of the  $2_m^+$  expectations is slightly lower than the generator level distributions suggest. The reason for this is that the efficiency  $\times$  acceptance ratio of the  $0^+$  to the  $2_m^+$  is not flat in this bin, as shown in Fig. 5.11, which distorts the shape contrary to what might be expected. It can be seen that the data is consistent with being flat. The  $\chi^2$  compatibility between the data and the SM expectation is  $p(\chi^2) = 86\%$ .

#### 7.4.2 Hypothesis tests of the SM Higgs, $0^+$ , vs. graviton-like, $2_m^+$

The separation between either of the two models and the data is extracted using the test statistic,  $q$ , defined as twice the negative ratio of the likelihoods for the  $0^+$  signal plus background hypothesis and the  $2_m^+$  signal plus background hypothesis when performing a simultaneous fit of all forty event classes together,  $q = -2 \ln(\mathcal{L}_{2_m^+ + \text{bkg.}} / \mathcal{L}_{0^+ + \text{bkg.}})$ .

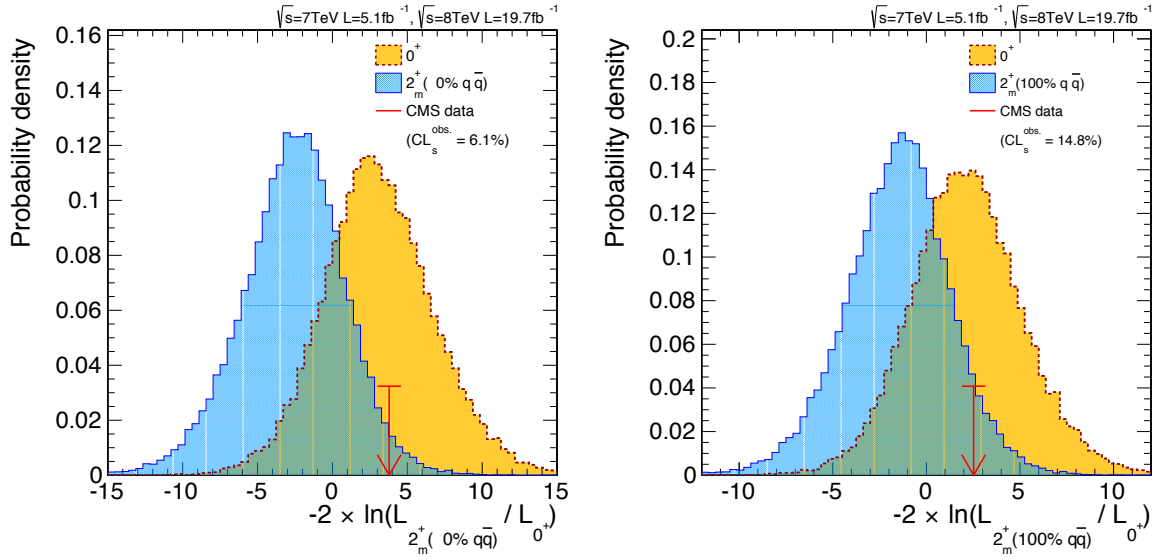




**Figure 7.15:** The SM extracted signal yield as a function of  $|\cos(\theta_{CS}^*)|$  for the  $0^+$  expectation (red line),  $2_m^+$  expectation with gluon fusion production only (blue line), the  $2_m^+$  expectation with quark-antiquark annihilation production only (green line), the  $2_m^+$  expectation with half  $gg$ , half  $q\bar{q}$  production (magenta line) and the observation (black points).

The distribution of this test statistic is shown in Fig. 7.16 for pseudo-experiments generated with an overall signal yield and signal position which is extracted from a fit to the data for the  $0^+$  hypothesis and the  $2_m^+$  hypothesis. The  $1 - CL_s$  observed exclusion for a gluon fusion-only produced spin-2 boson is 94% whilst for quark-antiquark produced boson is 85%.

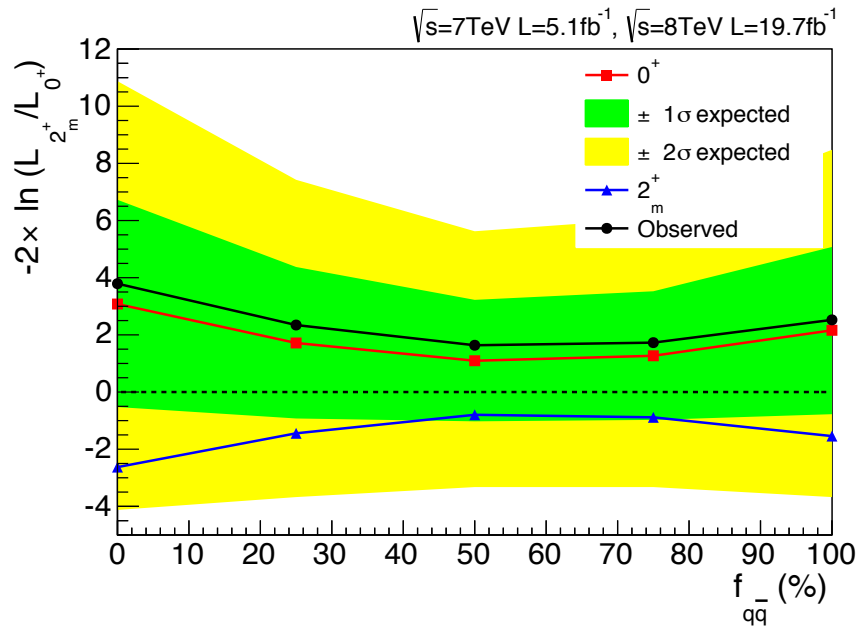
The previous two tests are both performed assuming that the  $2_m^+$  state is produced entirely by either gluon-fusion or quark-antiquark annihilation. A further three points, with mixtures of  $gg$  and  $q\bar{q}$  spin-2 production, of 25%, 50% and 75%, have been tested such that the overall yield of the  $2_m^+$  signal is fixed to the best fit value of the model in question to data and the fraction of  $q\bar{q}$  production is increased by a factor,  $f_{q\bar{q}}$ . Figure 7.17 shows the distribution of the test statistic as a function of the fraction of  $2_m^+$  production from  $q\bar{q}$  annihilation. Figure 7.16 is, in effect, a projection of Fig. 7.17 at the points  $f_{q\bar{q}} = 0\%$  and  $f_{q\bar{q}} = 100\%$ . It can be seen that the data is very much in line with the SM expectation. Whilst *a priori* it may look as though the data points in Fig. 7.17 lie too close to the SM mean (red line), all of these points are highly correlated. If the data



**Figure 7.16:** The distribution of the test statistic for pseudo experiments generated under the SM,  $0^+$ , hypothesis (orange) and the *graviton-like*,  $2_m^+$ , hypothesis (blue) with gluon fusion production only (left) and quark-antiquark production only (right). The observed value in the data is shown as the red arrow.

look flat in  $|\cos(\theta_{\text{CS}}^*)|$  then they will look flat for all values of  $f_{q\bar{q}}$ . In this sense the green and yellow bands in Fig. 7.17 can be misleading as they do not show these correlations.

These results show an independent discovery, in the  $H \rightarrow \gamma\gamma$  channel alone, of the Higgs like state around 125 GeV which was announced by the CMS and ATLAS collaborations in 2012[2, 1]. Furthermore, measurements of the properties of the observed resonance in this decay channel indicate a particle very consistent with the SM prediction.



**Figure 7.17:** The distribution of the test statistic for pseudo-experiments generated according to the SM,  $0^+$ , hypothesis (red) and the *graviton-like*,  $2_m^+$ , hypothesis (blue) as a function of the fraction of  $q\bar{q}$  production relative to  $gg$  production. The observed distribution in the data is shown by the black points.



# Chapter 8

## Conclusions

Results have been presented for a search of the SM Higgs boson decaying into two photons at the CMS experiment. An excess of events is observed over the background expectation with a local significance of  $5.7\sigma$ , where the SM expectation is  $5.2\sigma$ , constituting a standalone discovery of the particle first observed by the ATLAS and CMS experiments in 2012[1, 2]. The analysis strategy is to split events into a number of non-overlapping categories in order to increase the sensitivity to a signal and reduce the errors on measurements of the signal's couplings. The observed signal strength of the particle, relative to the SM Higgs boson prediction, is  $\sigma/\sigma_{SM} = 1.14^{+0.26}_{-0.23} \left[ {}^{+0.21}_{-0.21}(\text{stat})^{+0.13}_{-0.09}(\text{theory})^{+0.09}_{-0.05}(\text{syst}) \right]$ . The observed excess is more apparent in the 7 TeV dataset in which the signal strength is found to be  $\sigma/\sigma_{SM} = 2.22^{+0.60}_{-0.54}$ , compared to a value of  $\sigma/\sigma_{SM} = 0.90^{+0.25}_{-0.23}$  for the 8 TeV dataset. When forcing the signal in the two separate datasets to have the same mass the compatibility between these two measurements is  $1.9\sigma$ . The observed mass of the particle is  $m_H = 124.72 \pm 0.35 \left[ 0.31 (\text{stat}) \pm 0.16 (\text{syst}) \right]$  GeV. To probe the coupling of the observed state to fermions and bosons, relative to the SM, the signal strength is reparametrised to consist of two separate components, with a common mass, namely that from  $ggH$  and  $t\bar{t}H$  production,  $\mu_{ggH,t\bar{t}H}$ , and from  $VBF$  and  $VH$  production,  $\mu_{VBF,VH}$ . The observed values of these parameters is  $\mu_{ggH,t\bar{t}H} = 1.13^{+0.37}_{-0.31}$  and  $\mu_{VBF,VH} = 1.16^{+0.63}_{-0.57}$  showing a high level of compatibility with the SM prediction. The observed compatibility between the signal strength from each of the separate production modes,  $ggH$ ,  $VBF$ ,  $VH$  and  $t\bar{t}H$ , is found to have a probability,  $p(\chi^2) = 49\%$ . The observed compatibility between all of the analysis categories is found to have a probability,  $p(\chi^2) = 74\%$ . An analysis constructed to study the spin of the observed state is found to show consistency with the SM at a level of  $p(\chi^2) = 86\%$ . A spin-2 graviton produced entirely by gluon

fusion is excluded at 94% C.L. (where 92% is expected) and a spin-2 graviton produced entirely by quark-antiquark annihilation is excluded at 85% C.L. (where 83% is expected).

In summary there is a clear observation of the new state in this channel, and it is found to be very compatible with the SM Higgs. Furthermore, many studies of this particle's properties, since its discovery, in other decay modes and at other experiments suggest similar agreement with the SM prediction. Whilst there is strong theoretical motivation for the existence of a Higgs-like state somewhere in the low mass region (see Chapter 2) it is quite remarkable that many of the results shown here and elsewhere are compatible with predictions made several decades ago. Perhaps this is not the most interesting configuration of nature we could have hoped to see at the LHC. It is beyond doubt that there is physics beyond the SM; the Higgs self-coupling is quadratically divergent, neutrinos are not massless, the universe is known to consist of more than just the matter fermions discussed in Sec. 2.1 with a plethora of evidence for dark matter and dark energy, the lack of anti-matter in the universe does not fit with our expectation from creation and annihilation and the difference is not accounted for by current measurements of charge-parity violation, etc. There are theories which can explain some of these features, and many detectors looking for direct experimental signatures of them, but discussion of these is beyond the scope of this thesis. However, it is apparent that more detailed study of the Higgs sector could provide insight. Any deviations from the SM, probed by precise measurements of the observed boson's couplings, could provide evidence for new physics, whether through the loop in  $H \rightarrow \gamma\gamma$  decays or via decay chains such as  $X \rightarrow H \rightarrow Y$  for a new particle  $X$  and known decay products  $Y$ . Consequently, it is really only the beginning of Higgs physics. The discovery has happened and now the focus must turn to precision measurements of the particle's properties in the hope that it points the way to a new sector of unification in particle physics.

*Matthew Kenzie*

# Bibliography

- [1] ATLAS Collaboration, “Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC”, *Phys. Lett. B* **716** (2012) 1–29.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30–61.
- [3] F. Halzen and A. D. Martin, “Quarks and Leptons: An Introductory Course in Modern Particle Physics”. John Wiley and Sons, 1st edition, 1984.
- [4] J. Beringer et al., “Particle Data Group”, *Phys. Rev. D* **86** (2012) 010001.
- [5] E. Noether, “Invariante Variationsprobleme”, *Nachr. D. Konig. Gesellsh. D. Wiss. Zu Gottingen Math-phys. Klasse* (1918) 235–257.
- [6] E. Noether, “M. A. Tavel’s English translation of Invariante Variationsprobleme”, *Transport Theory and Statistical Physics* **1** (1971) 183–207.
- [7] M. W. Guidry, “Gauge Field Theories: An Introduction with Applications”. John Wiley and Sons, 1st edition, 1991.
- [8] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029/5582(61)90469-2.
- [9] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.
- [10] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc.* **C680519** 367–377.
- [11] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- [13] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.*

- 12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.
- [14] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [15] Y. Nambu, “Quasi-Particles and Gauge Invariance in the Theory of Superconductivity”, *Phys. Rev.* **117** (1960) 648–663, doi:10.1103/PhysRev.117.648.
- [16] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961) 154–164, doi:10.1007/BF02812722.
- [17] P. W. Anderson, “Plasmons, Gauge Invariance, and Mass”, *Phys. Rev.* **130** (Apr, 1963) 439–442, doi:10.1103/PhysRev.130.439.
- [18] G. ’t Hooft and M. Veltman, “Regularization and Renormalization of Gauge Fields”, *Nucl. Phys. B* **44** (1972) 189–213, doi:10.1016/0550-3213(72)90279-9.
- [19] G. Hooft, “Renormalizable Lagrangians for massive Yang-Mills fields”, *Nucl. Phys. B* **35** (1971), no. 1, 167 – 188, doi:http://dx.doi.org/10.1016/0550-3213(71)90139-8.
- [20] The UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at  $\sqrt{s} = 540$  GeV”, *Phys. Lett. B* **122** (1983), no. CERN-EP-83-13, 103–116. 31 p.
- [21] The UA1 Collaboration, “Experimental observation of lepton pairs of invariant mass around 95 GeV/ $c^2$  at the CERN SPS collider”, *Phys. Lett. B* **126** (1985), no. CERN-EP-83-073, 398–410. 17 p.
- [22] U. M. Heller, M. Klonfuss, H. Neuberger, and P. M. Vranas, “Numerical analysis of the Higgs mass triviality bound”, *Nucl. Phys. B* **405** (1993) 555–573, doi:10.1016/0550-3213(93)90559-8, arXiv:hep-ph/9303215.
- [23] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, “Search for the standard model Higgs boson at LEP”, *Phys. Lett. B* **565** (2003) 61–75, doi:10.1016/S0370-2693(03)00614-2, arXiv:hep-ex/0306033.
- [24] P. B. Renton, “Electroweak fits and constraints on the Higgs mass”, arXiv:hep-ph/0410177.
- [25] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, arXiv:1307.1347.
- [26] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [27] J. Anderson et al., “Snowmass Energy Frontier Simulations”, arXiv:1309.1057.
- [28] CMS Collaboration, “The CMS tracker system project: Technical Design Report”.



- Technical Design Report CMS. CERN, Geneva, 1997.
- [29] CMS Collaboration, “CMS tracking performance result from early LHC operation”, *Eur. Phys. J. C* **70** (2010) 1165–1192, doi:10.1140/epjc/s10052-010-1491-3, arXiv:hep-ex/1007.1988.
  - [30] CMS Collaboration, “CMS Tracker alignment and material budget measurement”, Technical Report CMS-CR-2011-164, CERN, Geneva, (Sep, 2011).
  - [31] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
  - [32] CMS Collaboration, “2012 ECAL detector performance plots”, *CMS-DP-2013-007* (2012).
  - [33] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and  $E_T^{miss}$ ”, *CMS-PFT-09-001* (2009).
  - [34] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, “Electron reconstruction in the CMS Electromagnetic Calorimeter”, *CMS-AN-2001-034* (2001).
  - [35] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9.
  - [36] J. Weng, “A Global Event Description using Particle Flow with the CMS Detector”, arXiv:hep-ex/0810.3686v1.
  - [37] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm”, *J. High Energy Phys.* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063.
  - [38] C. Ferro, “B-tagging in CMS”, arXiv:hep-ex/1201.5292.
  - [39] CMS Collaboration, “Jet energy corrections and uncertainties in CMS: reducing their impact on physical measurements”, *J. Phys. Conf. Ser.* **404** (2012) 012014.
  - [40] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659** (2008), no. 1–2, 119 – 126, doi:http://dx.doi.org/10.1016/j.physletb.2007.09.077.
  - [41] Hastie, T. and Tibshirani, R. and Friedman, J., “The Elements of Statistical Learning”. Springer-Verlag, 2nd edition, 2009.
  - [42] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.”, *The Annals of Statistics* **29** (10, 2001) 1189–1232, doi:10.1214/aos/1013203451.
  - [43] J. H. Friedman, “Stochastic gradient boosting”, *Computational Statistics and Data Analysis* **38** (2002), no. 4, 367 – 378, doi:http://dx.doi.org/10.1016/S0167-9473(01)00065-2. Nonlinear Methods and Data Mining.

- [44] A. Hoecker et al., “TMVA - Toolkit for Multivariate Data Analysis”, [arXiv:physics/0703039](#).
- [45] F. J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit”, *JSTOR* **46** (1951) 68–78.
- [46] CMS Collaboration, “Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update”, *CMS-PAS-SMP-12-008* (2012).
- [47] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting: Summer 2013 Update”, *CMS-PAS-LUM-13-001* (2013).
- [48] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO Higgs boson production via gluon fusion matched with shower in POWHEG”, *JHEP* **04** (2009) 002, [doi:10.1088/1126-6708/2009/04/002](#).
- [49] P. Nason and C. Oleari, “NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG”, *JHEP* **02** (2010) 037, [doi:10.1007/JHEP02\(2010\)037](#).
- [50] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **0605** (2006) 026, [doi:10.1088/1126-6708/2006/05/026](#).
- [51] Y. Gao et al., “Spin determination of single-produced resonances at hadron colliders”, *Phys. Rev. D* **81** (2010) 242–245.
- [52] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* **0902** (2009) 007, [doi:10.1088/1126-6708/2009/02/007](#), [arXiv:0811.4622](#).
- [53] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **1106** (2011) 128, [doi:10.1007/JHEP06\(2011\)128](#), [arXiv:1106.0522](#).
- [54] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, [doi:10.1016/S0168-9002\(03\)01368-8](#).
- [55] E. Chabanat and N. Estre, “Deterministic Annealing for Vertex Finding at CMS”,.
- [56] T. Skwarnicki, “A Study of the Radiative Cascade Transitions between the upsilon-prime and upsilon resonances”, *DESY-F31-86-02*.
- [57] CMS Collaboration, “Measurement of the inclusive W and Z production cross sections in pp collisions at  $\sqrt{s}=7$  TeV with the CMS experiment”, *JHEP* **2011** (2011) 1, [doi:10.1007/JHEP10\(2011\)132](#).
- [58] R. Szalapski and D. Zeppenfeld, “Probing color singlet exchange in Z + two jet events at the CERN LHC”, *Phys. Rev. D* **54** (1996) 6680–6689, [doi:10.1103/PhysRevD.54.6680](#), [arXiv:arXiv:hep-ph/9605444](#).
- [59] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **06** (2011) P11002, [doi:10.1088/1748-0221/6/11/P11002](#).

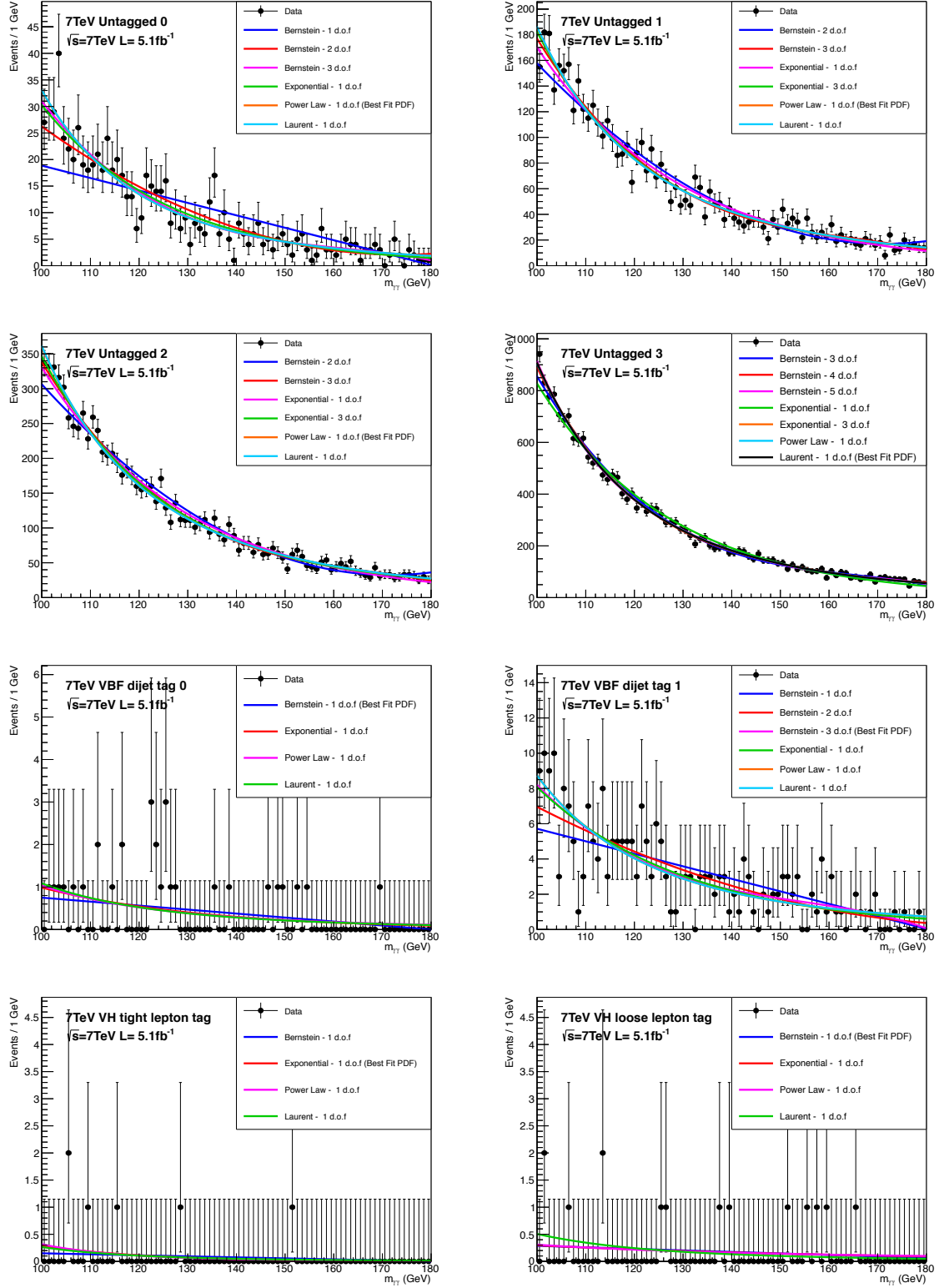
- [60] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett.* **B659** (2008) 119, doi:10.1016/j.physletb.2007.09.077.
- [61] M. Cacciari, G. P. Salam, and G. Soyez, “The Catchment Area of Jets”, *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005.
- [62] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, arXiv:1111.6097.
- [63] CMS Collaboration, “MET performance in 8 TeV data”, *CMS PAS* **CMS-PAS-JME-12-002** (2012).
- [64] W. S. Cleveland, “Robust locally weighted regression and smoothing scatterplots”, *Journal of the American Statistical Association* **74** (1979), no. 368, 829–836.
- [65] L. D. Landau, “On the angular momentum of a two-photon system”, *Dokl. Akad. Nauk. Ser. Fiz.* **60** (1948) 207.
- [66] C. N. Yang, “Selection Rules for the Dematerialization of a Particle Into Two Photons”, *Phys. Rev.* **77** (1950) 242.
- [67] Y. Gao et al., “Spin determination of single-produced resonances at hadron colliders”, *Phys. Rev. D* **81** (2010) 242–245.
- [68] J. C. Collins and D. E. Soper, “Angular distribution of dileptons in high-energy hadron collisions”, *Phys. Rev. D.* **16** (1977) 2219–2225.
- [69] F. James, “Statistical Methods in Experimental Physics”. World Scientific, 2nd edition, 2006.
- [70] H. Bozdogan, “Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions”, *Psychometrika* **52** (1987), no. 3, 345–370, doi:10.1007/BF02294361.
- [71] G. M. Phillips, “On generalized Bernstein polynomials”, *Numerical Analysis: A. R. Mitchell 75th Birthday Volume* (1996) 263–269.
- [72] G. M. Phillips, “Bernstein polynomials based on the q-integers”, *The heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T.J. Rivlin, Ann. Numer. Math.* **4** (1997) 511–518.
- [73] R. A. Fisher, “On the Interpretation of  $\chi^2$  from Contingency Tables, and the Calculation of P”, *Journal of the Royal Statistical Society* **85** (1922), no. 1, 87–94.
- [74] J. H. Friedman, “Multivariate Adaptive Regression Splines”, *The Annals of Statistics* **19** (03, 1991) 1–67, doi:10.1214/aos/1176347963.
- [75] G. R. Terrell and D. W. Scott, “Variable Kernel Density Estimation”, *The Annals of Statistics* **20** (1992), no. 3, 1236–1265.
- [76] I. W. Stewart and F. J. Tackmann, “Theory Uncertainties for Higgs and Other

- Searches Using Jet Bins”, *Phys. Rev. D* **85** (2012) 034011, doi:10.1103/PhysRevD.85.034011, arXiv:1107.2117.
- [77] R. Eusebi, “Jet energy corrections and uncertainties in CMS: reducing their impact on physics measurements”, *Journal of Physics: Conference Series* **404** (2012).
- [78] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1–19.
- [79] A. L. Read, “Presentation of search results: the CLs technique”, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002) 2693.

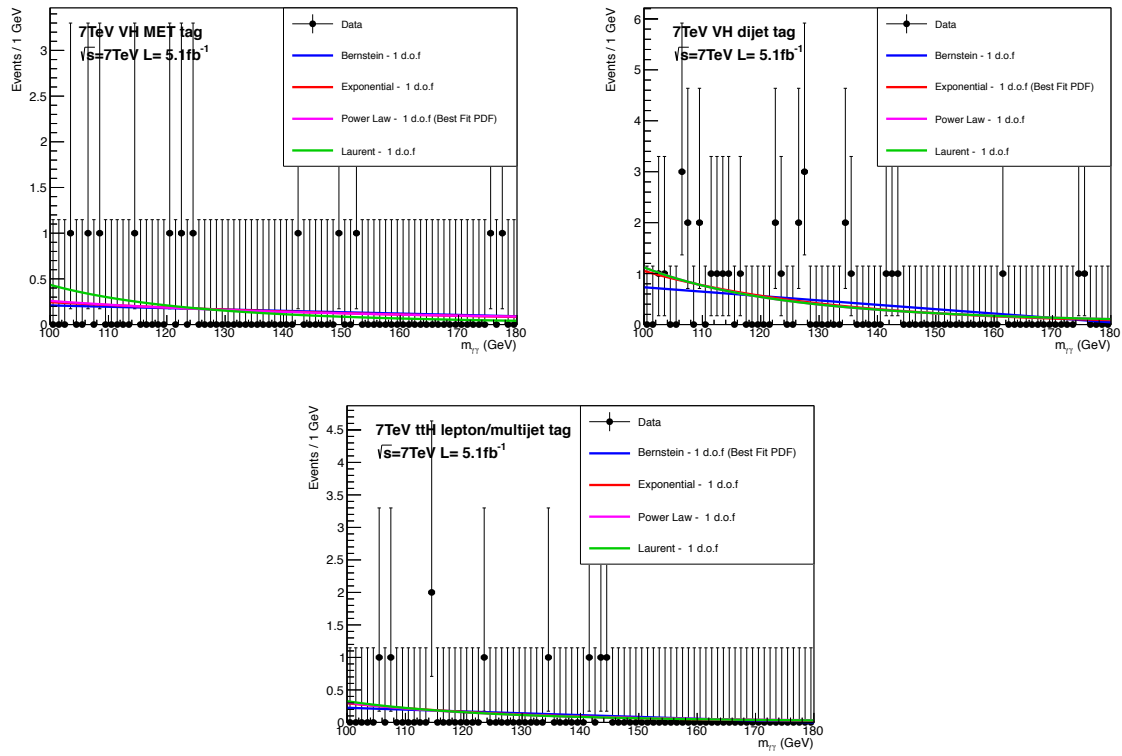
# Appendix A

## Envelope function choices

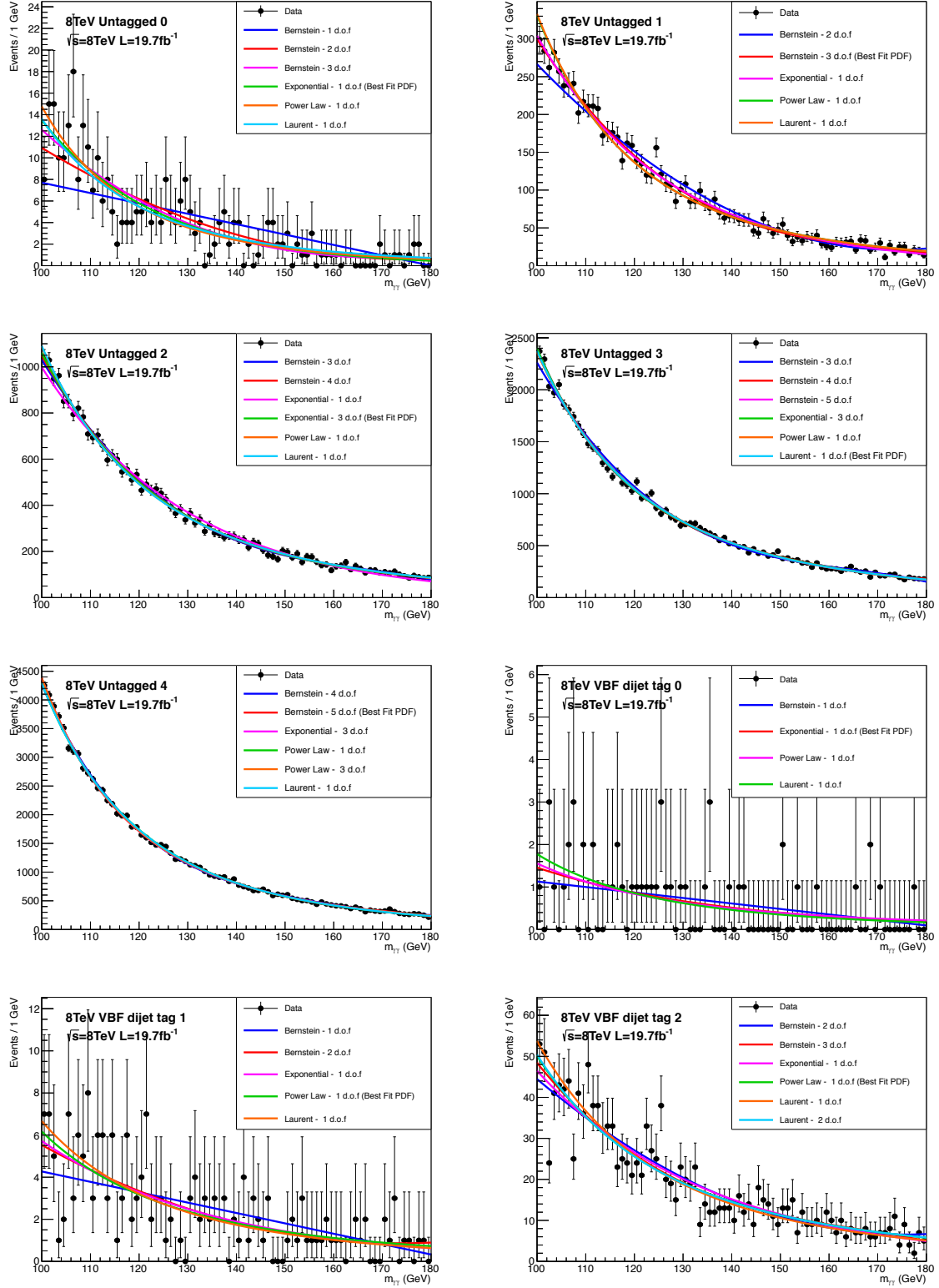
The invariant mass distributions with the different envelope functions used, after having been fit to the data, for each of the analysis categories in the MFM analysis.



**Figure A.1:** The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the inclusive, dijet and  $VH$  lepton tag categories in the 7 TeV dataset.

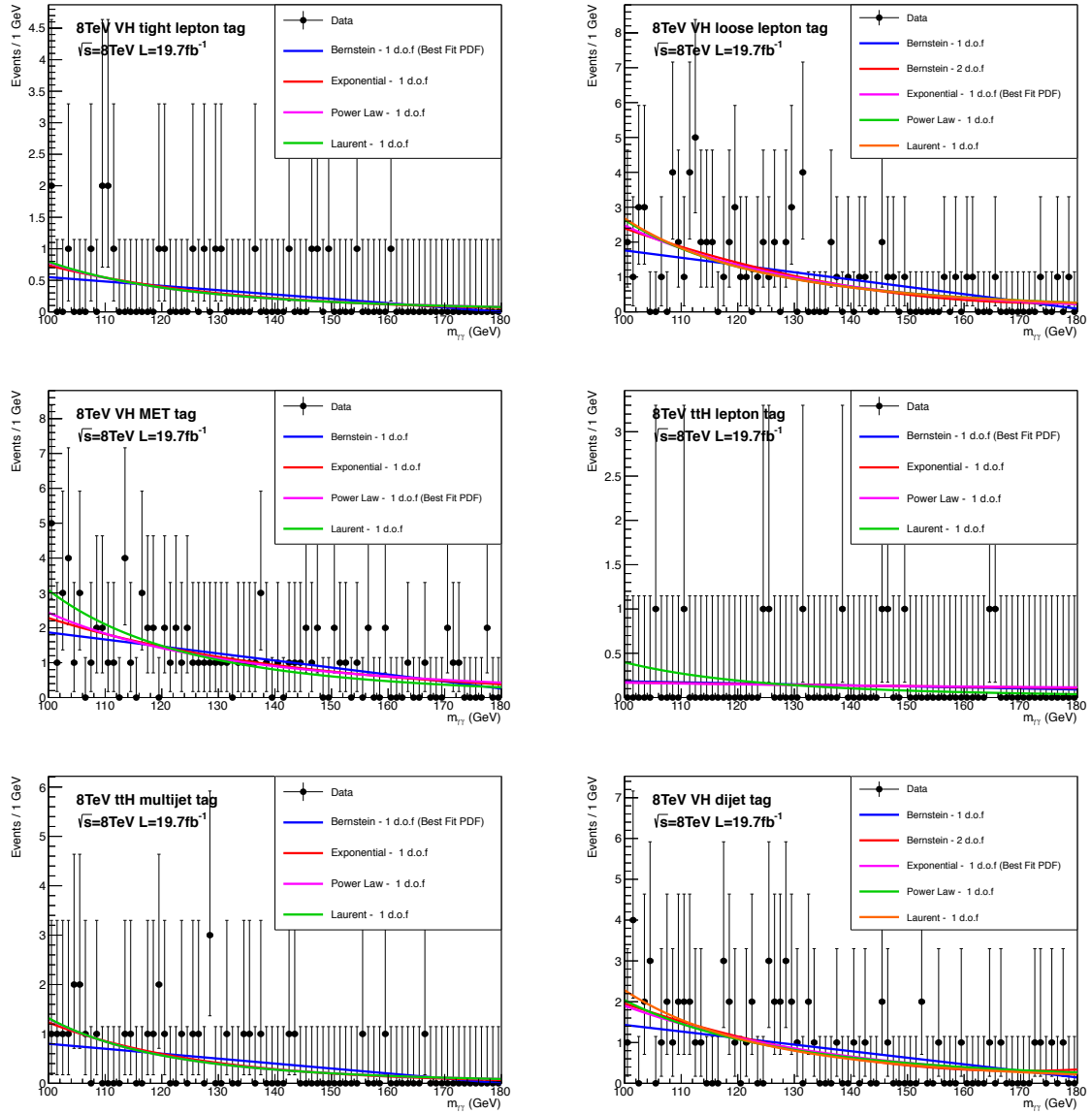


**Figure A.2:** The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the  $VH \cancel{E}_T$  and jet tag and  $t\bar{t}H$  categories in the 7 TeV dataset.



**Figure A.3:** The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the inclusive and  $VBF$  dijet tag categories in the 8 TeV dataset.





**Figure A.4:** The diphoton invariant mass distribution and the background function choices profiled using the envelope method for the  $VH$  and  $t\bar{t}H$  tag categories in the 8 TeV dataset.