Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector


DOI 10.1007/JHEP01(2015)069

Publication date 2015

Document Version Final published version

Published in The Journal of High Energy Physics

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A search for the $b\bar{b}$ decay of the Standard Model Higgs boson is performed with the ATLAS experiment using the full dataset recorded at the LHC in Run 1. The integrated luminosities used are 4.7 and 20.3 fb$^{-1}$ from $pp$ collisions at $\sqrt{s} = 7$ and 8 TeV, respectively. The processes considered are associated $(W/Z)H$ production, where $W \rightarrow e\nu/\mu\nu$, $Z \rightarrow ee/\mu\mu$ and $Z \rightarrow \nu\nu$. The observed (expected) deviation from the background-only hypothesis corresponds to a significance of 1.4 (2.6) standard deviations and the ratio of the measured signal yield to the Standard Model expectation is found to be $\mu = 0.52 \pm 0.32\text{(stat.)} \pm 0.24\text{(syst.)}$ for a Higgs boson mass of 125.36 GeV. The analysis procedure is validated by a measurement of the yield of $(W/Z)Z$ production with $Z \rightarrow b\bar{b}$ in the same final states as for the Higgs boson search, from which the ratio of the observed signal yield to the Standard Model expectation is found to be $0.74 \pm 0.09\text{(stat.)} \pm 0.14\text{(syst.)}$.

KEYWORDS: Hadron-Hadron Scattering

ArXiv ePrint: 1409.6212
## Contents

1 Introduction 1

2 The ATLAS detector 3

3 Data and simulated samples 3

4 Object reconstruction 6

5 Event selection 8

6 Multivariate analysis 14

7 Background composition and modelling 21
   7.1 Multijet background 21
   7.2 Corrections to the simulation 23
   7.3 Distributions in the dijet-mass analysis 25
   7.4 Distributions in the multivariate analysis 29

8 Systematic uncertainties 34
   8.1 Experimental uncertainties 34
   8.2 Uncertainties on the multijet backgrounds 35
   8.3 Uncertainties on the modelling of the simulated backgrounds 36
   8.4 Uncertainties on the signal modelling 40

9 Statistical procedure 42
   9.1 General aspects 42
   9.2 Technical details 44
   9.3 Cross checks using diboson production 46

10 Analysis of the 7 TeV data 50
   10.1 Object reconstruction 50
   10.2 Event selection 50
   10.3 Background composition and modelling 50
   10.4 Systematic uncertainties 51
   10.5 Statistical procedure 51

11 Results 53
   11.1 Nominal results 53
   11.2 Cross-check with the dijet-mass analysis 55
   11.3 Cross-check with the diboson analysis 58

12 Summary 60
1 Introduction

For decades, the Higgs boson [1–4] of the Standard Model (SM) remained an unconfirmed prediction. In July 2012, the ATLAS and CMS experiments at the LHC reported the observation of a new particle with a mass of about 125 GeV and with properties consistent with those expected for the SM Higgs boson [5, 6]. Since then, more precise measurements have strengthened the hypothesis that the new particle is indeed a Higgs boson [7–9]. These measurements, however, have been mainly performed in the bosonic decay modes of the new particle (\(H \rightarrow \gamma\gamma, H \rightarrow ZZ,\) and \(H \rightarrow WW\)). It is also essential to verify whether it decays into fermions as predicted by the Standard Model. Recently, the CMS Collaboration reported evidence for the \(\tau\tau\) decay mode of the Higgs boson at a level of significance of 3.4 standard deviations (\(\sigma\)) for \(m_H = 125\) GeV [10].

The \(H \rightarrow b\bar{b}\) decay mode is predicted in the SM to have a branching ratio of 58% for \(m_H = 125\) GeV [11]. Accessing \(H \rightarrow b\bar{b}\) decays is therefore crucial for constraining, under fairly general assumptions [12], the overall Higgs boson decay width and, in a global fit to all accessible combinations of Higgs boson production and decay modes, to allow for measurements of absolute Higgs boson couplings. An inclusive search for \(H \rightarrow b\bar{b}\) is not feasible at hadron colliders because of the overwhelming background from multijet production. In spite of a cross section more than an order of magnitude lower than the dominant gluon-fusion process, associated production of a Higgs boson with a vector boson, \(W\) or \(Z\) [13], offers a viable alternative because leptonic decays of the vector boson, \(W \rightarrow \ell\nu, Z \rightarrow \ell\ell (\ell = e,\mu),\) and \(Z \rightarrow \nu\nu\), can be efficiently used for triggering and background reduction purposes [14, 15]. The CDF and D0 experiments at the Tevatron reported an excess of events in their search for associated \((W/Z)H\) production in the \(H \rightarrow b\bar{b}\) decay mode at a significance level of 2.8\(\sigma\) for \(m_H = 125\) GeV [16]. Recently, the CMS experiment reported an excess of events in the \(H \rightarrow b\bar{b}\) decay mode with a significance of 2.1\(\sigma\) for \(m_H = 125\) GeV [17].

In this paper, a search for associated \((W/Z)H\) production of the SM Higgs boson in the \(b\bar{b}\) decay mode is presented, using the full integrated luminosity accumulated by ATLAS during Run 1 of the LHC: 4.7 and 20.3 \(\text{fb}^{-1}\) from proton-proton (\(pp\)) collisions at centre-of-mass energies of 7 and 8 TeV in 2011 and 2012, respectively. An analysis of the 7 TeV dataset has already been published by ATLAS [18]. In addition to the increase in the amount of data analysed, the update presented in this paper benefits from numerous analysis improvements. Some of the improvements to the object reconstruction, however, are available only for the 8 TeV dataset, which leads to separate analysis strategies for the two datasets.
The analysis is performed for events containing zero, one, or two charged leptons (electrons or muons), targeting the $Z \to \nu\nu$, $W \to \ell\nu$, or $Z \to \ell\ell$ decay modes of the vector boson, respectively. In addition to $Z \to \nu\nu$ decays, the 0-lepton channel has a smaller but not insignificant contribution from leptonic $W$ decays when the lepton is produced outside of the detector acceptance or not identified. A $b$-tagging algorithm is used to identify the jets consistent with originating from an $H \to b\bar{b}$ decay. To improve the sensitivity, the three channels are each split according to the vector-boson transverse momentum, the number of jets (two or three), and the number of $b$-tagged jets. Topological and kinematic selection criteria are applied within each of the resulting categories.

A binned maximum likelihood fit is used to extract the signal yield and the background normalisations. Systematic uncertainties on the signal and background modelling are implemented as deviations in their respective models in the form of “nuisance” parameters that are varied in the fit. Each nuisance parameter is constrained by a penalty term in the likelihood, associated with its uncertainty. Two versions of the analysis are presented in this paper: in the first, referred to as the dijet-mass analysis, the mass of the dijet system of $b$-tagged jets is the final discriminating variable used in the statistical analysis; in the other, a multivariate analysis (MVA) incorporating various kinematic variables in addition to the dijet mass, as well as $b$-tagging information, provides the final discriminating variable. Because the latter information is not available in similar detail for the 7 TeV dataset, the MVA is used only for the 8 TeV dataset. In both analyses, dedicated control samples, typically with loosened $b$-tagging requirements, constrain the contributions of the dominant background processes. The most significant background sources are $(W/Z)+$heavy-flavour-jet production and $t\bar{t}$ production. The normalisations of these backgrounds are fully determined by the likelihood fit. Other significant background sources are single-top-quark and diboson ($WZ$ and $ZZ$) production, with normalisations taken from theory, as well as multijet events, normalised using multijet-enriched control samples. Since the MVA has higher expected sensitivity, it is chosen as the nominal analysis for the 8 TeV dataset to extract the final results. To validate the analysis procedures, both for the dijet-mass and MVA approaches, a measurement of the yield of $(W/Z)Z$ production is performed in the same final states and with the same event selection, with $H \to b\bar{b}$ replaced by $Z \to b\bar{b}$.

This paper is organised as follows. A brief description of the ATLAS detector is given in section 2. Details of the data and simulated samples used in this analysis are provided in section 3. This is followed by sections describing the dijet-mass and multivariate analyses applied to the 8 TeV data. The reconstruction of physics objects such as leptons and jets is addressed in section 4. Section 5 details the event selections applied to the dijet-mass and multivariate analyses, while section 6 explains the construction of the final discriminating variable of the MVA. Section 7 discusses the background composition in the various analysis regions, while the systematic uncertainties are addressed in section 8. The statistical procedure used to extract the results is described in section 9. For the 7 TeV data, only a dijet-mass analysis is used, and differences with respect to the 8 TeV data analysis are specified in section 10. The results are presented and discussed in section 11, and a summary of the paper is given in section 12.
2 The ATLAS detector

The ATLAS detector [19] is cylindrically symmetric around the beam axis and is structured in a barrel and two endcaps. It consists of three main subsystems. The inner tracking detector is immersed in the 2 T axial magnetic field produced by a superconducting solenoid. Charged-particle position and momentum measurements are made by pixel detectors followed by silicon-strip detectors in the pseudorapidity\(^1\) range \(|\eta| < 2.5\) and by a straw-tube transition-radiation tracker (TRT) in the range \(|\eta| < 2.0\). The pixel detectors are crucial for \(b\)-tagging, and the TRT also contributes to electron identification. The calorimeters, located beyond the solenoid, cover the range \(|\eta| < 4.9\) with a variety of detector technologies. The liquid-argon electromagnetic calorimeters are divided into barrel (\(1.475 < |\eta| < 3.2\)), endcap (\(1.375 < |\eta| < 3.2\)), and forward (\(3.1 < |\eta| < 4.9\)) sections. The hadronic calorimeters (using scintillator tiles or liquid argon as active materials) surround the electromagnetic calorimeters with a coverage of \(|\eta| < 4.9\). The muon spectrometer measures the deflection of muon tracks in the field of three large air-core toroidal magnets, each containing eight superconducting coils. It is instrumented with separate trigger and high-precision tracking chambers covering the \(|\eta| < 2.4\) and \(|\eta| < 2.7\) ranges, respectively.

The trigger system is organised in three levels. The first level is based on custom-made hardware and uses coarse-granularity calorimeter and muon information. The second and third levels are implemented as software algorithms and use the full detector granularity. At the second level, only regions deemed interesting at the first level are analysed, while the third level, called the event filter, makes use of the full detector read-out to reconstruct and select events, which are then logged for offline analysis at a rate of up to 400 Hz averaged over an accelerator fill.

3 Data and simulated samples

The datasets used in this analysis include only \(pp\) collision data recorded in stable beam conditions and with all relevant sub-detectors providing high-quality data. The corresponding integrated luminosities are 4.7 and 20.3 fb\(^{-1}\) [20] for the 7 and 8 TeV data, respectively.

Events in the 0-lepton channel are selected by triggers based on the magnitude \(E_T^{\text{miss}}\) of the missing transverse momentum vector. The \(E_T^{\text{miss}}\) trigger configuration evolved during data taking to cope with the increasing luminosity, and the trigger efficiency was improved for the 8 TeV data. The dependence of the \(E_T^{\text{miss}}\) trigger efficiency on the \(E_T^{\text{miss}}\) reconstructed offline is measured in \(W \rightarrow \mu\nu+\text{jets}\) and \(Z \rightarrow \mu\mu+\text{jets}\) events collected with single-muon triggers, with the offline \(E_T^{\text{miss}}\) calculated without the muon contribution. As there was a brief period of data-taking in which the \(E_T^{\text{miss}}\) triggers were not available for

---

\(^1\)ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the \(z\)-axis coinciding with the axis of the beam pipe. The \(x\)-axis points from the IP towards the centre of the LHC ring, and the \(y\)-axis points upward. Cylindrical coordinates \((r,\phi)\) are used in the transverse plane, \(\phi\) being the azimuthal angle around the \(z\)-axis. The pseudorapidity is defined in terms of the polar angle \(\theta\) as \(\eta = -\ln\tan(\theta/2)\). The distance in \((\eta,\phi)\) coordinates, \(\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}\), is also used to define cone sizes. Transverse momentum and energy are defined as \(p_T = p \sin \theta\) and \(E_T = E \sin \theta\), respectively. For the purpose of object selections, \(\eta\) is calculated relative to the geometric centre of the detector; otherwise, it is relative to the reconstructed primary vertex of each event.
the first bunch crossings of two bunch trains, the integrated luminosity for the 0-lepton channel in the 7 TeV dataset is reduced to 4.6 fb\(^{-1}\). Events in the 1-lepton channel are primarily selected by single-lepton triggers. The \(E_T\) threshold of the single-electron trigger was raised from 20 to 22 GeV during the 7 TeV data-taking period, and to 24 GeV for the 8 TeV data. The \(p_T\) threshold of the single-muon trigger was similarly increased from 18 GeV for the 7 TeV data to 24 GeV at 8 TeV. As the single-lepton triggers for the 8 TeV data include isolation criteria, triggers with higher thresholds (60 GeV for electrons and 36 GeV for muons) but no isolation requirements are used in addition. Single-lepton trigger efficiencies are measured using a tag-and-probe method applied to \(Z \rightarrow ee\) and \(Z \rightarrow \mu\mu\) events. In the 1-muon sub-channel, \(E_T^{miss}\) triggers are also used to compensate for the limited muon trigger-chamber coverage in some regions of the detector. Events in the 2-lepton channel are selected by a combination of single-lepton, dielectron and dimuon triggers. The thresholds of the dilepton triggers are 12 GeV for electrons and 13 GeV for muons.

Monte Carlo (MC) simulated samples are produced for signal and background processes using the \textsc{atlfast-II} simulation \cite{21}, which includes a full simulation of the ATLAS detector based on the \textsc{geant4} program \cite{22}, except for the response of the calorimeters for which a parameterised simulation is used. A list of the generators used for signal and background simulations is given in table 1.

The MC generator used for \(q\bar{q}\)-initiated \(WH\) and \(ZH\) production is \textsc{pythia8} \cite{23} with the CTEQ6L1 \cite{24} parton distribution functions (PDFs). The AU2 tune \cite{25, 26} is used for the parton shower, hadronisation, and multiple parton interactions. The \textsc{photos} program \cite{27} is used for QED final-state radiation. The \textsc{powheg} generator \cite{28, 29, 30} is used within the MiNLO approach \cite{31} with the CT10 PDFs \cite{32}, interfaced to \textsc{pythia8} with the AU2 tune, as a cross-check and to evaluate systematic uncertainties on the signal acceptance and kinematic properties. It is also used for the generation of gluon-gluon-initiated \(ZH\) production at leading order (LO) in QCD, with results cross-checked by an independent computation \cite{33}. (For the analysis of the 7 TeV data, the \textsc{pythia8} generator is used for \(gg \rightarrow ZH\).) The transverse momentum distributions of the Higgs boson show substantial differences between the two \(ZH\) production processes. For \(q\bar{q}\)-initiated \(WH\) and \(ZH\) production, the total production cross sections and associated uncertainties are computed at next-to-next-to-leading order (NNLO) in QCD \cite{34, 35, 36}, and with electroweak corrections at next-to-leading order (NLO) \cite{37}. Additional normalisation-preserving differential electroweak NLO corrections are applied as a function of the transverse momentum of the vector boson \cite{38}. For gluon-gluon-initiated \(ZH\) production, NLO corrections \cite{39}, which increase the total \(ZH\) production cross section by about 5%, are taken into account. The Higgs boson decay branching ratios are calculated with \textsc{hdecay} \cite{11}. Signal samples are simulated for Higgs boson masses from 100 to 150 GeV in steps of 5 GeV. All charged-lepton flavours are simulated in the \(W\) and \(Z\) decays, as leptonic decays of the \(\tau\) leptons can also be selected in the analysis. For the Higgs boson, only the \(b\bar{b}\) decay mode is considered in the analysis.

The main background processes are \((W/Z)+\text{jets}\) and \(t\bar{t}\) production. Version 1.4.1 of the \textsc{sherpa} generator \cite{40} is used with the CT10 PDFs to simulate \(W+\text{jets}\) and \(Z+\text{jets}\) at leading-order in QCD, with massive \(c\)- and \(b\)-quarks. For \(t\bar{t}\) production, the simulation is
Table 1. The generators used for the simulation of the signal and background processes. (⋆) For the analysis of the 7 TeV data, PYTHIA8 is used for the simulation of the $gg \to ZH$ process, and HERWIG for the simulation of diboson processes.

<table>
<thead>
<tr>
<th>Process</th>
<th>Generator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal((⋆))</td>
<td></td>
</tr>
<tr>
<td>$q\bar{q} \to ZH \to \nu\nu bb/\ell\ell bb$</td>
<td>PYTHIA8</td>
</tr>
<tr>
<td>$gg \to ZH \to \nu\nu bb/\ell\ell bb$</td>
<td>POWHEG+PYTHIA8</td>
</tr>
<tr>
<td>$q\bar{q} \to WH \to \ell\nu bb$</td>
<td>PYTHIA8</td>
</tr>
<tr>
<td>Vector boson + jets</td>
<td></td>
</tr>
<tr>
<td>$W \to \ell\nu$</td>
<td>SHERPA 1.4.1</td>
</tr>
<tr>
<td>$Z/\gamma^* \to \ell\ell$</td>
<td>SHERPA 1.4.1</td>
</tr>
<tr>
<td>$Z \to \nu\nu$</td>
<td>SHERPA 1.4.1</td>
</tr>
<tr>
<td>Top-quark</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>POWHEG+PYTHIA</td>
</tr>
<tr>
<td>$t$-channel</td>
<td>ACERMC+PYTHIA</td>
</tr>
<tr>
<td>$s$-channel</td>
<td>POWHEG+PYTHIA</td>
</tr>
<tr>
<td>$Wt$</td>
<td>POWHEG+PYTHIA</td>
</tr>
<tr>
<td>Diboson((⋆))</td>
<td></td>
</tr>
<tr>
<td>$WW$</td>
<td>POWHEG+PYTHIA8</td>
</tr>
<tr>
<td>$WZ$</td>
<td>POWHEG+PYTHIA8</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>POWHEG+PYTHIA8</td>
</tr>
</tbody>
</table>

performed with the POWHEG generator with the CT10 PDFs, interfaced with PYTHIA6 [41], for which the CTEQ6L1 PDFs and the Perugia2011C tune [25, 26] are used. In this analysis, the final normalisations of these dominant backgrounds are constrained by the data, but theoretical cross sections are used to optimise the selection. The cross sections are calculated at NNLO for ($W/Z$)+jets [42] and at NNLO, including resummations of next-to-next-to-leading logarithmic (NNLL) soft gluon terms, for $t\bar{t}$ [43].

Additional backgrounds arise from single-top-quark and diboson ($WW$, $WZ$, and $ZZ$) production. For single-top-quark production, the $s$-channel exchange process and $Wt$ production are simulated with POWHEG, as for $t\bar{t}$, while the $t$-channel exchange process is simulated with the ACERMC generator [44] interfaced with PYTHIA6, using the CTEQ6L1 PDFs and the Perugia2011C tune. The cross sections are taken from refs. [45–47]. The POWHEG generator with the CT10 PDFs, interfaced to PYTHIA8 with the AU2 tune, is used for diboson processes [48]. (For the analysis of the 7 TeV data, the HERWIG generator [49] is used instead with the CTEQ6L1 PDFs and the AUET2 tune [25, 26], and the cross sections are obtained at NLO from MCFM [50] with the MSTW2008NLO PDFs [51].)

Events from minimum-bias interactions are simulated with the PYTHIA8 generator with the MSTW2008LO PDFs [52] and the A2 tune [25, 26]. They are overlaid on the simulated
signal and background events according to the luminosity profile of the recorded data. The contributions from these “pile-up” interactions are simulated both within the same bunch crossing as the hard-scattering process and in neighbouring bunch crossings. The resulting events are then processed through the same reconstruction programs as the data.

Additional generators are used for the assessment of systematic uncertainties as explained in section 8. Simulated jets are labelled according to which generated hadrons with $p_T > 5$ GeV are found within a cone of size $\Delta R = 0.4$ around the reconstructed jet axis. If a $b$-hadron is found, the jet is labelled as a $b$-jet. If not and a $c$-hadron is found, the jet is labelled as a $c$-jet. If neither a $b$- nor a $c$-hadron is found, the jet is labelled as a light (i.e., $u$, $d$, or $s$-quark, or gluon) jet. Simulated $V$+jet events, where $V$ stands for $W$ or $Z$, are then categorised according to the labels of the two jets that are used to reconstruct the Higgs boson candidate. If one of those jets is labelled as a $b$-jet, the event belongs to the $Vb$ category. If not and one of the jets is labelled as a $c$-jet, the event belongs to the $Vc$ category. Otherwise, the event belongs to the $Vl$ category. Further subdivisions are defined according to the flavour of the other jet from the pair, using the same precedence order: $Vbb$, $Vbc$, $Vbl$, $Vcc$, $Vcl$. The combination of $Vbb$, $Vbc$, $Vbl$ and $Vcc$ is denoted $V$+hf. The $Vcl$ final state is not included in $V$+hf because the main production process is $gs \rightarrow Wc$ rather than gluon splitting.

4 Object reconstruction

In this section, the reconstruction of physics objects used in the analysis of the 8 TeV data is presented. Differences relevant for the analysis of the 7 TeV data are reported in section 10.

Charged-particle tracks are reconstructed with a $p_T$ threshold of 400 MeV. The primary vertex is selected from amongst all reconstructed vertices as the one with the largest sum of associated-track squared transverse momenta $\Sigma p_T^2$ and is required to have at least three associated tracks.

Three categories of electrons [53, 54] and muons [55] are used in the analysis, referred to as loose, medium and tight leptons in order of increasing purity. Loose leptons are selected with transverse energy $E_T > 7$ GeV. Loose electrons are required to have $|\eta| < 2.47$ and to fulfil the “very loose likelihood” identification criteria defined in ref. [54]. The likelihood-based electron identification combines shower-shape information, track-quality criteria, the matching quality between the track and its associated energy cluster in the calorimeter (direction and momentum/energy), TRT information and a criterion to help identify electrons originating from photon conversions. The electron energies are calibrated by making use of reference processes such as $Z \rightarrow ee$ [56]. Three types of muons are included in the loose definition to maximise the acceptance: (1) muons reconstructed in both the muon spectrometer and the inner detector (ID); (2) muons with $p_T > 20$ GeV identified in the calorimeter and associated with an ID track with $|\eta| < 0.1$, where there is limited muon-chamber coverage; and (3) muons with $|\eta| > 2.5$ identified in the muon spectrometer, and which do not match full ID tracks due to the limited inner-detector coverage. For muons of
the first and second type, the muon-track impact parameters with respect to the primary vertex must be smaller than 0.1 mm and 10 mm in the transverse plane and along the z-axis, respectively. Finally, the scalar sum of the transverse momenta of tracks within a cone of size $\Delta R = 0.2$ centred on the lepton-candidate track, excluding the lepton track, is required to be less than 10% of the transverse momentum of the lepton.

Medium leptons must meet the loose identification criteria and have $E_T > 25 \text{ GeV}$. Medium muons must be reconstructed in both the muon spectrometer and the inner detector and have $|\eta| < 2.5$. Tight electrons are required to additionally fulfil the “very tight likelihood” identification criteria [54]. For both the tight electrons and the tight muons, more stringent isolation criteria must be satisfied: the sum of the calorimeter energy deposits in a cone of size $\Delta R = 0.3$ around the lepton, excluding energy associated with the lepton candidate, must be less than 4% of the lepton energy, and the track-based isolation requirement is tightened from 10% to 4%.

Jets are reconstructed from noise-suppressed topological clusters of energy in the calorimeters [57] using the anti-$k_t$ algorithm [58] with a radius parameter of 0.4. Jet energies are corrected for the contribution of pile-up interactions using a jet-area-based technique [59] and calibrated using $p_T$- and $\eta$-dependent correction factors determined from simulation, with residual corrections from in situ measurements applied to data [60, 61]. Further adjustments are made based on jet internal properties, which improve the energy resolution without changing the average calibration (global sequential calibration [60]). To reduce the contamination by jets from pile-up interactions, the scalar sum of the $p_T$ of tracks matched to the jet and originating from the primary vertex must be at least 50% of the scalar sum of the $p_T$ of all tracks matched to the jet. This requirement is only applied to jets with $p_T < 50 \text{ GeV}$ and $|\eta| < 2.4$. Jets without any matched track are retained. The jets kept for the analysis must have $p_T > 20 \text{ GeV}$ and $|\eta| < 4.5$.

To avoid double-counting, the following procedure is applied to loose leptons and jets. First, if a jet and an electron are separated by $\Delta R < 0.4$, the jet is discarded. Next, if a jet and a muon are separated by $\Delta R < 0.4$, the jet is discarded if it has three or fewer matched tracks since in this case it is likely to originate from a muon having showered in the calorimeter; otherwise the muon is discarded. (Such muons are nevertheless included in the computation of the $E_{\text{miss}}^T$ and in the jet energy corrections described in section 5.) Finally, if an electron and a muon are separated by $\Delta R < 0.2$, the muon is kept unless it is identified only in the calorimeter, in which case the electron is kept.

The MV1c $b$-tagging algorithm is used to identify jets originating from $b$-quark fragmentation. This algorithm combines in a neural network the information from various algorithms based on track impact-parameter significance or explicit reconstruction of $b$- and $c$-hadron decay vertices. It is an improved version of the MV1 algorithm [62–64] with higher $c$-jet rejection. Four $b$-tagging selection criteria (or operating points) are calibrated and used in the analysis, corresponding to average efficiencies of 80%, 70%, 60% and 50% for $b$-jets with $p_T > 20 \text{ GeV}$, as measured in simulated $t\bar{t}$ events. In this analysis, the 80%, 70% and 50% operating points are denoted loose, medium and tight, respectively. For the tight (loose) operating point, the rejection factors are 26 (3) and 1400 (30) against $c$-jets and light jets, respectively. For the tight operating point, the $c$-jet rejection factor is 1.9 times larger than obtained with the MV1 algorithm.
The $b$-tagging efficiencies for $b$-jets, $c$-jets and light jets are measured in both data and simulation using dedicated event samples such as $t\bar{t}$ events for $b$-jets, events with identified $D^*$ mesons for $c$-jets, or multijet events for light jets. The small differences observed are used to correct the simulation by so-called “scale factors” (SFs) within intervals between two operating points. These SFs are parameterised as a function of the jet $p_T$ and, for light jets, also $|\eta|$. The SFs are, however, strictly valid only for the generator used to derive them. The differences observed when the efficiencies are measured with different generators are taken into account by additional “MC-to-MC” SFs. Such differences can be caused by, e.g., different production fractions of heavy-flavour hadrons or modelling of their decays.

Because of the large cross sections of $Vl$ and $Vc$ production, these backgrounds remain significant despite the powerful rejection of non-$b$-jets by the $b$-tagging algorithm. It is impractical to simulate a sufficiently large number of $Vl$ and $Vc$ events to provide a reliable description of these backgrounds in the analysis samples for which two $b$-tagged jets are required. An alternative procedure, parameterised tagging, is therefore used. Here, instead of directly tagging the $c$- and $l$-labelled jets with the MV1c algorithm, parameterisations as functions of $p_T$ and $|\eta|$ of their probabilities to be $b$-tagged are used for the $Vl$, $Vc$ and $WW$ processes in all analysis samples in which two $b$-tagged jets are required. These parameterisations are, however, integrated over other variables that can affect the $c$- and light-jet tagging efficiencies. In particular, a strong dependence of these efficiencies is observed on $\Delta R$, the angular separation from the closest other jet, and a significant difference is seen between direct and parameterised tagging for $Vcc$ events with $\Delta R < 1$. No such difference is seen for $Vcl$, $Vl$ and $WW$ events. A dedicated correction, depending on $\Delta R$, is therefore applied to the $Vcc$ events.

The missing transverse momentum vector $E_{\text{miss}}^{\text{T}}$ [65, 66] is measured as the negative vector sum of the transverse momenta associated with energy clusters in the calorimeters with $|\eta| < 4.9$. Corrections are applied to the energies of clusters associated with reconstructed objects (jets, electrons, $\tau$ leptons, and photons), using the calibrations of these objects. The transverse momenta of reconstructed muons are included, with the energy deposited by these muons in the calorimeters properly removed to avoid double-counting. In addition, a track-based missing transverse momentum vector, $p_{\text{T}}^{\text{miss}}$, is calculated as the negative vector sum of the transverse momenta of tracks with $|\eta| < 2.4$ associated with the primary vertex.

Additional corrections are applied to the simulation to account for small differences from data for trigger efficiencies, for lepton reconstruction and identification efficiencies, as well as for lepton energy and momentum resolutions.

5 Event selection

In this section, the event selection applied in the analysis of the 8 TeV data is presented. Differences in the analysis of the 7 TeV data are reported in section 10.

The analysis is optimised for a Higgs boson mass of 125 GeV. Events are first categorised according to the numbers of leptons, jets, and $b$-tagged jets.
Events containing no loose leptons are assigned to the 0-lepton channel. Events containing one tight lepton and no additional loose leptons are assigned to the 1-lepton channel. Events containing one medium lepton and one additional loose lepton of the same flavour, and no other loose leptons, are assigned to the 2-lepton channel. In the 1- and 2-lepton channels, for at least one of the lepton triggers by which the event was selected, the objects that satisfied the trigger are required to be associated with the selected leptons.

The jets used in this analysis, called “selected jets”, must have $p_T > 20$ GeV and $|\eta| < 2.5$, the $\eta$ range within which $b$-tagging can be applied. There must be exactly two or three such selected jets. Events containing an additional jet with $p_T > 30$ GeV and $|\eta| > 2.5$ are discarded to reduce the $t\bar{t}$ background. Only selected jets are considered further, e.g., to define the jet multiplicity, or to calculate kinematic variables.

The $b$-tagging algorithm is applied to all selected jets. There must be no more than two such jets loosely $b$-tagged, and 3-jet events in which the lowest-$p_T$ jet is loosely $b$-tagged are discarded. At least one of the two $b$-tagged jets must have $p_T > 45$ GeV. The following $b$-tagging categories are then defined as shown in figure 1. Events with two jets satisfying the tight $b$-tagging criterion form the TT (or Tight) category; those not classified as TT, but with two jets satisfying the medium $b$-tagging criterion, form the MM (or Medium) category; those not classified as TT or MM, but with two jets satisfying the loose $b$-tagging criterion, form the LL (or Loose) category. This categorisation improves the sensitivity with respect to what would be obtained using a single category, such as TT+MM, with the LL category providing constraints on the backgrounds not containing two real $b$-jets. Events with exactly one jet loosely $b$-tagged form the 1-tag category, and those with no loosely $b$-tagged jet form the 0-tag category. In the 3-jet categories, the dijet system is formed by the two $b$-tagged jets in any of the 2-tag categories, by the $b$-tagged jet and the leading (highest-$p_T$) non-$b$-tagged jet for events in the 1-tag category, and by the two leading jets in the 0-tag category.

Additional topological and kinematic criteria are applied to reject background events and enhance the sensitivity of the search. They are outlined in table 2 and detailed below. In general, the selection criteria are looser in the MVA than in the dijet-mass analysis in order to maximise the information available to the final discriminant.

Further categorisation is performed according to the transverse momentum of the vector boson, $p_T^V$, to take advantage of the better signal-to-background ratio at high $p_T^V$. The transverse momentum of the vector boson is reconstructed as the $E_T^{\text{miss}}$ in the 0-lepton channel, the magnitude $p_T^V$ of the vector sum of the lepton transverse momentum and the $E_T^{\text{miss}}$ in the 1-lepton channel, and the magnitude $p_T^Z$ of the vector sum of the transverse momenta of the two leptons in the 2-lepton channel. In the dijet-mass analysis, the events are categorised in five $p_T^V$ intervals, with boundaries at 0, 90, 120, 160, and 200 GeV. In the 0-lepton channel and for events fulfilling the condition on $\sum p_T^{\text{jet}i}$ mentioned in table 2, the $E_T^{\text{miss}}$ trigger is fully efficient for $E_T^{\text{miss}} > 160$ GeV, 97% efficient for $E_T^{\text{miss}} = 120$ GeV, and 80% efficient for $E_T^{\text{miss}} = 100$ GeV, with an efficiency that decreases rapidly for lower $E_T^{\text{miss}}$. Only four intervals are therefore used in the 0-lepton channel, with a minimum $E_T^{\text{miss}}$ value of 100 GeV. In the 1-muon sub-channel, the $E_T^{\text{miss}}$ trigger is used for $p_T^V > 120$ GeV to recover events not selected by the single-muon trigger, thus increasing the signal acceptance.
Figure 1. Event classification as a function of the output of the MV1c b-tagging algorithm for the two highest $p_T$ jets. The bin boundaries denote the operating points (MV1c(jet) OP) as defined in section 4, corresponding to b-tagging efficiencies of 100%, 80%, 70%, 50%, i.e., the b-jet purity increases from left (bottom) to right (top). The event categories are 0-tag, 1-tag, and TT, MM and LL for 2-tag, as explained in the text.

in this channel by 8%. In the MVA, only two intervals are defined, with $p_T^V$ below or above 120 GeV, but the detailed $p_T^V$ information is used in the final discriminant.

In the dijet-mass analysis, requirements are applied to the angular separation between the two jets of the dijet system, $\Delta R(jet_1, jet_2)$, which depend on the $p_T^V$ interval. The requirement on the minimum value reduces the background from $V$+jet production, while the requirement on the maximum value, which reduces the background from $t\bar{t}$ production, is tightened with increasing $p_T^V$ to take advantage of the increasing collimation of the dijet system for the signal. To increase the signal acceptance, the requirement on the minimum value is removed in the highest $p_T^V$ interval, where the amount of background is smallest.

In the MVA, where the $\Delta R(jet_1, jet_2)$ information is used in the final discriminant, only a minimum value is required, a requirement which is also removed for $p_T^V > 200$ GeV.

In the 0-lepton channel, the multijet (MJ) background is suppressed by imposing requirements on the magnitude $p_T^{miss}$ of the track-based missing transverse momentum vector $p_T^{miss}$, the azimuthal angle between $E_T^{miss}$ and $p_T^{miss}$, $\Delta \phi(E_T^{miss}, p_T^{miss})$, the azimuthal angle between $E_T^{miss}$ and the nearest jet, min[$\Delta \phi(E_T^{miss}, jet)$], and the azimuthal angle between $E_T^{miss}$ and the dijet system, $\Delta \phi(E_T^{miss}, dijet)$. In addition, a minimum value is required for the scalar sum of the jet transverse momenta, $\sum p_T^{jet}$, which depends on the jet multiplicity. Additional requirements are applied in the lowest $p_T^V$ interval of the 0-lepton channel, where the MJ background is largest: $N_{jet} = 2$; $E_T^{miss} > 100$ GeV; $\Delta \phi(jet_1, jet_2) < 2.7$; $S > 7$; and $L > 0.5$. Here, $\Delta \phi(jet_1, jet_2)$ is the azimuthal angle be-
Variable | Dijet-mass analysis | Multivariate analysis
---|---|---
\(p_T^V\) [GeV] | 0–90 | 90(∗)–120 | 120–160 | 160–200 | > 200 | 0–120 | > 120
\(\Delta R(jet_1, jet_2)\) | 0.7–3.4 | 0.7–3.0 | 0.7–2.3 | 0.7–1.8 | < 1.4 | > 0.7 (\(p_T^V<200\) GeV)

<table>
<thead>
<tr>
<th>0-lepton selection</th>
</tr>
</thead>
</table>
\(p_T^{\text{miss}}\) [GeV] | > 30 | > 30 | | | > 30 | < \(\pi/2\) | < \(\pi/2\)
\(\Delta \phi(E_T^{\text{miss}}, p_T^{\text{miss}})\) | | | | | | | |
\(\text{min}[\Delta \phi(E_T^{\text{miss}}, \text{jet})]\) | NU | NU | | | | | |
\(\Delta \phi(E_T^{\text{miss}}, \text{dijet})\) | > 2.2 | > 2.8 | | | | | |
\(\sum_{i=1}^{N_{\text{jet}}=2(3)} p_{T_i}^{\text{jet}}\) [GeV] | > 120 (NU) | > 120 (150) | | | > 120 (150) |

<table>
<thead>
<tr>
<th>1-lepton selection</th>
</tr>
</thead>
</table>
\(m_W^T\) [GeV] | | | | | | | |
\(H_T\) [GeV] | > 180 | | | | | | |
\(E_T^{\text{miss}}\) [GeV] | | | > 20 | | > 50 | – | > 20 |

<table>
<thead>
<tr>
<th>2-lepton selection</th>
</tr>
</thead>
</table>
\(m_{\ell\ell}\) [GeV] | 83-99 | 71-121 | | | | | |
\(E_T^{\text{miss}}\) [GeV] | < 60 | | | | | | |

**Table 2.** Event topological and kinematic selections. NU stands for ‘Not Used’. (∗) In the 0-lepton channel, the lower edge of the second \(p_T^V\) interval is set at 100 GeV instead of 90 GeV. For the 1-lepton channel, only the 1-muon sub-channel is used in the \(p_T^V<120\) GeV intervals.

tween the two jets, \(S\) is the \(E_T^{\text{miss}}\) significance, defined as the ratio of \(E_T^{\text{miss}}\) to the square root of \(\sum p_{T_i}^{\text{jet}}\); and \(L\) is a likelihood ratio constructed to discriminate further against the MJ background.\(^2\)

In the 1-lepton channel, a requirement is imposed on the transverse mass\(^3\) \(m_W^T\) in the dijet-mass analysis. This requirement reduces the contamination from the \(t\bar{t}\) background. Requirements are also imposed on \(H_T\) (\(E_T^{\text{miss}}\)) for \(p_T^V<120\) GeV, where \(H_T\) is the scalar sum of \(E_T^{\text{miss}}\) and the transverse momenta of the two leading jets and the lepton. This mainly reduces the MJ background. As discussed in section 7.1, the MJ background is difficult to model and remains substantial in the 1-electron sub-channel in the \(p_T^V<120\) GeV intervals. Therefore, only the 1-muon sub-channel is used in these intervals.

In the 2-lepton channel, criteria are imposed on the dilepton invariant mass, \(m_{\ell\ell}\), which

\(^2\)The likelihood ratio uses the following inputs: \(\Delta \phi(E_T^{\text{miss}}, \text{dijet}); \Delta \phi(jet_1, jet_2);\) the magnitude of the vector sum of the two jet transverse momenta, \(H_T^{\text{miss}}, H_T^{\text{miss}}\) divided by \(\sum p_{T_i}^{\text{jet}}\); and the cosine of the helicity angle in the dijet rest frame as defined in ref. \([67]\). For the MJ background, the probability density functions used in the likelihood ratio are constructed from data events selected with \(\Delta \phi(E_T^{\text{miss}}, p_T^{\text{miss}})>\pi/2\).

\(^3\)The transverse mass \(m_W^T\) is calculated from the transverse momentum and the azimuthal angle of the charged lepton, \(p_T^\ell\) and \(\phi^\ell\), and from the missing transverse momentum magnitude, \(E_T^{\text{miss}}\), and azimuthal angle, \(\phi^{\text{miss}}\): \(m_W^T = \sqrt{2 p_T^\ell E_T^{\text{miss}} (1 - \cos(\phi^\ell - \phi^{\text{miss}}))}\).
must be consistent with the mass of the $Z$ boson. In the dijet-mass analysis a requirement is imposed on $E_{T}^{\text{miss}}$; this variable is used in the final discriminant of the MVA.

For events in which two jets are loosely $b$-tagged, these selection criteria define a set of “2-tag signal regions”, categorised in terms of channel (0, 1, or 2 leptons), $p_T$ interval, and jet multiplicity (2 or 3). In the dijet-mass analysis, a further division is performed into the TT, MM and LL $b$-tagging categories. In the MVA, where the $b$-tagging information is used in the final discriminant, a similar subdivision is performed with the difference that the TT and MM categories are merged in the 0- and 2-lepton channels. Similarly defined 1-tag and 0-tag “control regions” are used in the analysis to constrain the main backgrounds. In the 1-lepton channel, the 2-tag signal regions with a third selected jet act in practice as control regions because they are largely dominated by $t\bar{t}$ events. All 2-tag signal and 1-tag control regions are used simultaneously in the global fit (described in section 9) used to extract the results. The 0-tag control regions are used only for background modelling studies (reported in section 7).

After event selection, the energy calibration of the $b$-tagged jets is improved as follows. The energy from muons within a jet is added to the calorimeter-based jet energy after removing the energy deposited by the muon in the calorimeter (muon-in-jet correction), and a $p_T$-dependent correction is applied to account for biases in the response due to resolution effects (resolution correction). This latter correction is determined for the $p_T$ spectrum of jets from the decay of a Higgs boson with $m_H = 125$ GeV in simulated $(W/Z)H$ events. The dijet mass resolution for the signal is improved by 14% after these corrections and is typically 11% (figure 2(a)). In the 2-lepton channel, wherein there is no true $E_{T}^{\text{miss}}$ involved except possibly from semileptonic heavy-flavour decays, the energy calibration of the jets is further improved by a kinematic likelihood fit, which includes a Breit-Wigner constraint on the dilepton mass, Gaussian constraints on each of the transverse components of the $\ell\ell bb$ system momentum (with a width of 9 GeV, as determined from $ZH$ simulated events), dedicated transfer functions relating the true jet transverse momenta to their reconstructed values (after the muon-in-jet correction, but without the resolution correction) as well as a prior built from the expected true jet $p_T$ spectrum in $ZH$ events (playing a role similar to the resolution correction). Overall, the $b\bar{b}$ mass resolution is improved by 30% in the 2-lepton channel (figure 2(b)).

The cross sections times branching ratios for $(W/Z)H$ with $W \rightarrow \ell\nu, Z \rightarrow \ell\ell, Z \rightarrow \nu\nu,$ and $H \rightarrow b\bar{b}$, as well as the acceptances in the three channels after full selection are given in table 3 for the MVA and the dijet-mass analysis. The acceptance for other production and decay modes of the Higgs boson is negligible. The 0-lepton channel adds 7% in acceptance for the $W \rightarrow \ell\nu$ process with respect to the 1-lepton channel. Similarly, the 1-lepton channel adds 10% in acceptance for the $Z \rightarrow \ell\ell$ process with respect to the 2-lepton channel.
**Figure 2.** Dijet-invariant-mass distribution for the decay products of a Higgs boson with \( m_H = 125 \text{ GeV} \) in the 2-lepton MVA selection. The distributions are shown (a) using jets after global sequential calibration (GSC, solid), and after adding muons inside jets (dotted) and after correcting for resolution effects specific to the kinematics of the decay of a Higgs boson with \( m_H = 125 \text{ GeV} \) (dash-dotted); (b) using jets after global sequential calibration (GSC, solid), and after adding muons inside jets and applying the kinematic fit (dash-dotted). The distributions are fit to the Bukin function [68] and the parameter representing the width of the core of the distribution is shown in the figures, as well as the relative improvement in the resolution with respect to jets after the global sequential calibration.

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Process} & \text{Cross section} \times \text{BR [fb]} & \text{Acceptance [%]} \\
& & \text{0-lepton} & \text{1-lepton} & \text{2-lepton} \\
\hline
q\bar{q} \rightarrow (Z \rightarrow \ell\ell)(H \rightarrow b\bar{b}) & 14.9 & \text{--} & 1.3 (1.1) & 13.4 (10.9) \\
\text{gg} \rightarrow (Z \rightarrow \ell\ell)(H \rightarrow b\bar{b}) & 1.3 & \text{--} & 0.9 (0.7) & 10.5 (8.1) \\
q\bar{q} \rightarrow (W \rightarrow \ell\nu)(H \rightarrow b\bar{b}) & 131.7 & 0.3 (0.3) & 4.2 (3.7) & \text{--} \\
q\bar{q} \rightarrow (Z \rightarrow \nu\nu)(H \rightarrow b\bar{b}) & 44.2 & 4.0 (3.8) & \text{--} & \text{--} \\
\text{gg} \rightarrow (Z \rightarrow \nu\nu)(H \rightarrow b\bar{b}) & 3.8 & 5.5 (5.0) & \text{--} & \text{--} \\
\hline
\end{array}
\]

**Table 3.** The cross section times branching ratio (BR) and acceptance for the three channels at 8 TeV. For \( ZH \), the \( q\bar{q} \) - and \( gg \)-initiated processes are shown separately. The branching ratios are calculated considering only decays to muons and electrons for \( Z \rightarrow \ell\ell \), decays to all three lepton flavours for \( W \rightarrow \ell\nu \) and decays to neutrinos for \( Z \rightarrow \nu\nu \). The acceptance is calculated as the fraction of events remaining in the combined 2-tag signal regions of the MVA (dijet-mass analysis) after the full event selection.
6 Multivariate analysis

Although the dijet mass is the kinematic variable that provides the best discrimination between signal and backgrounds, the sensitivity of the search is improved by making use of additional kinematic, topological and $b$-tagging properties of the selected events in a multivariate analysis. The Boosted Decision Tree (BDT) technique \cite{69, 70} is used, which, similarly to other multivariate methods, properly accounts for correlations between variables.

Dedicated BDTs are constructed, trained and evaluated in each of the 0-, 1- and 2-lepton channels in the 2-tag regions (with the LL, MM and TT categories combined) and separately for the events with two and three jets. In the 0-lepton channel, only events with $p_T^V > 120\text{ GeV}$ are used, whereas for the 1- and 2-lepton channels individual BDTs are used for $p_T^V < 120\text{ GeV}$ and $p_T^V > 120\text{ GeV}$. Events in the electron and muon sub-channels are combined since none of the variables used are lepton-flavour specific. In the 0-lepton channel, the final results are obtained using the MVA for $p_T^V > 120\text{ GeV}$. For the small $100 < p_T^V < 120\text{ GeV}$ interval, which has reduced sensitivity, no dedicated BDT is trained and only the dijet-mass distribution is used.

The BDTs are trained to separate the $(VH, H \rightarrow b\bar{b})$ signal from the sum of the expected background processes. The input variables used to construct the BDTs are chosen in order to maximise the separation, while avoiding the use of variables not improving the performance significantly. Starting from the dijet mass, additional variables are tried one at a time and the one yielding the best separation gain is kept. This procedure is repeated until adding more variables does not result in a significant performance gain. The final sets of variables for the different channels are listed in table 4. The $b$-tagged jets belonging to the dijet system (with mass denoted $m_{bb}$) are labelled in decreasing $p_T$ as $b_1$ and $b_2$, and their separation in pseudorapidity is $|\Delta\eta(b_1, b_2)|$. The $b$-tagging information is provided by the outputs of the MV1c neural network, $MV1c(b_1)$ and $MV1c(b_2)$. The angular separation, in the transverse plane, of the vector boson and the dijet system of $b$-tagged jets and their pseudorapidity separation are denoted $\Delta\phi(V, bb)$ and $|\Delta\eta(V, bb)|$, respectively. In the 0-lepton channel, $H_T$ is defined as the scalar sum of the transverse momenta of all jets and $E_T^{\text{miss}}$. In the 1-lepton channel, the angle between the lepton and the closest $b$-tagged jet in the transverse plane is denoted $\min[\Delta\phi(\ell, b)]$. The other variables were defined in the previous sections. In 3-jet events, the third jet is labelled as jet$_3$ and the mass of the 3-jet system is denoted $m_{bbj}$.

The input variables of the BDTs are compared between data and simulation, and good agreement is found within the assessed uncertainties. Selected input-variable distributions are shown in figure 3.\textsuperscript{4} In this figure, as for all figures in this section, the MJ background is estimated as described in section 7.1, corrections to the simulation as explained in section 7.2 are applied, and background normalisations and shapes are adjusted by the global fit of the MVA as outlined at the beginning of section 7 and presented in more detail in section 9. A similarly good agreement is found for the correlations between pairs of input variables, as can be seen in figure 4.

\textsuperscript{4}In this and all similar figures, all backgrounds are taken into account, but those contributing less than 1% are omitted from the legend.
Table 4. Variables used in the multivariate analysis for the 0-, 1- and 2-lepton channels.

<table>
<thead>
<tr>
<th>Variable</th>
<th>0-Lepton</th>
<th>1-Lepton</th>
<th>2-Lepton</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T^V$</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>$E_T^{\text{miss}}$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$p_T^{b_1}$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$b_1$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$m_{\text{bb}}$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$\Delta R(b_1, b_2)$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \eta(b_1, b_2)</td>
<td>$</td>
<td>×</td>
</tr>
<tr>
<td>$\Delta \phi(V, bb)$</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\Delta \eta(V, bb)</td>
<td>$</td>
<td></td>
</tr>
<tr>
<td>$H_T$</td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>$\min[\Delta \phi(\ell, b)]$</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{W}$</td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>$m_{\text{eff}}$</td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>$MV1c(b_1)$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>$MV1c(b_2)$</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Only in 3-jet events

The Toolkit for Multivariate Data Analysis, TMVA [71], is used to train the BDTs. The values for the training parameters are found by determining the configuration with the best separation between signal and background in a coarsely binned multi-dimensional training parameter space, followed by more finely grained one-dimensional scans of individual training parameters. In order to make use of the complete set of simulated MC events for the BDT training and evaluation in an unbiased way, the MC events are split into two samples of equal size, $A$ and $B$. The performance of the BDTs trained on sample $A$ ($B$) is evaluated with sample $B$ ($A$) in order to avoid using identical events for both training and evaluation of the same BDT. Half of the data are analysed with the BDTs trained on sample $A$, and the other half with the BDTs trained on sample $B$. At the end, the output distributions of the BDTs trained on samples $A$ and $B$ are merged for both the simulated and data events.

The values of the BDT outputs do not have a well-defined interpretation. A dedicated procedure is applied to transform the BDT-output distributions to obtain a smoother distribution for the background processes and a finer binning in the regions with the largest signal contribution, while at the same time preserving a sufficiently large number of background events in each bin. Starting from a very fine-binned histogram of the BDT-output distribution, the procedure merges histogram bins, from high to low BDT-output values, until a certain requirement, based on the fractions of signal and background events in the merged bin, is satisfied. To limit the number of bins and to reduce the impact of statistical fluctuations, a further condition is that the statistical uncertainty of the expected
Figure 3. Examples of variables input to the BDT in the 2-jet 2-tag category (LL, MM and TT combined) for $p_T^{V} > 120$ GeV: (a) 0-lepton channel, dijet mass; (b) 0-lepton channel, $E_T^{miss}$; (c) 1-lepton channel, $\Delta R(b_1,b_2)$; (d) 1-lepton channel, $p_T^{V}$; (e) 2-lepton channel, $p_T^{b_1b_2}$; (f) 2-lepton channel, $\Delta \eta(V,b\bar{b})$. The distributions for the 2-lepton channel in (e) and (f) are shown after having applied the kinematic fit as described in section 5. The background contributions after the global fit of the MVA are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The entries in overflow are included in the last bin. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 4. Correlations between the dijet mass and $\Delta R(b_1, b_2)$ input variables to the BDT in the 2-jet 2-tag category (LL, MM and TT combined) of the 0-lepton channel for $p_T > 120$ GeV: (a) dijet mass versus $\Delta R(b_1, b_2)$ for the total expected background (shadings indicating the numbers of events) and the data (open boxes with the box size being proportional to the number of events), (b) and (c) show the mean values and RMS of the projections onto the $\Delta R(b_1, b_2)$ and dijet-mass axes, respectively, for the total expected background after the global fit of the MVA and the data.
total background contribution has to be smaller than 10\% in each merged bin. The free parameters of the transformation algorithm are optimised to maximise the expected signal sensitivity. For simplicity, these transformed outputs, which are used for the analysis, are called “BDT$_{VH}$ discriminants” in the following. An optimisation of the number of bins and bin boundaries is also performed for the $m_{bb}$ distribution used in the dijet-mass analysis in a similar way, where the free parameters of the transformation algorithm are optimised separately for the different analysis regions. The effect of the transformation on the BDT-output and dijet-mass distributions can be seen in figure 5 for the 1-lepton channel and one signal region. The transformation groups into few bins the $m_{bb}$ regions that are far from the signal on each of the low and high mass sides, while it expands the region close to the signal mass, where the signal-to-background ratio is largest. The effect on the BDT output is similar, but simpler to visualise because the signal and the background accumulate initially on the high and the low sides of the distribution, respectively.

Correlations between input variables and the BDT$_{VH}$ discriminant can provide information on the impact of individual variables on the classification. Figure 6 shows such correlations for the dijet mass, which is the BDT input that provides the best single-variable discriminating power.
Figure 5. Top: the dijet-mass distributions for the expected background and signal contributions in the 1-lepton channel and the 2-jet 2-tag TT category for 160 GeV < \( p_T^W \) ≤ 200 GeV (a) before and (b) after applying the transformation of the histogram bins. Bottom: the BDT-output distribution for the expected background and signal contributions in the 1-lepton channel and the 2-jet 2-tag TT category for \( p_T^W > 120 \) GeV (c) before and (d) after applying the transformation of the histogram bins. The background contributions after the relevant global fit (of the dijet-mass analysis in (a) and (b) and of the MVA in (c) and (d)) are shown as filled histograms. The Higgs boson signal (\( m_H = 125 \) GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as \( \mu = 1.0 \)), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The entries in overflow are included in the last bin. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 6. Correlations between the dijet mass and the BDT, $\nu$ discriminant in the 2-jet 2-tag category (LL, MM and TT combined) of the 0-lepton channel for $p_T > 120$ GeV: (a) the dijet mass versus the BDT, $\nu$ discriminant for the total expected background (shadings indicating the numbers of events) and the data (open boxes with the box size being proportional to the number of events), (b) and (c) show the mean values and RMS of the projections onto the axes of the BDT, $\nu$ discriminant and dijet mass, respectively, for the total expected background after the global fit of the MVA and the data.
7 Background composition and modelling

This section describes the modelling of individual backgrounds. In many cases, the data are able to constrain the normalisations and shapes better than the a priori estimates. A likelihood fit (also called “global fit”) is used to simultaneously extract both the signal yield and constraints on the background normalisations and shapes. The distributions used by the fit are those of the dijet mass or BDT_{VH} discriminant in the 2-tag signal regions, as appropriate, as well as those of the MV1c value of the b-tagged jet in the 1-tag control regions. More details are provided in section 9.

For the multijet (MJ) backgrounds, the normalisations and shapes provided as inputs to the fit are estimated from data, as explained below. For the other backgrounds the inputs are taken from the simulation, except for the normalisations of the V+jets and tt̄ backgrounds that are left free to float in the fit. The corrections to these two backgrounds, described below, are applied prior to the fit.

In all distributions presented in this section, unless otherwise specified, the normalisations of the various backgrounds are those extracted from the global fit for the dijet-mass or multivariate analysis, as appropriate. The fit also adjusts the background shapes in those distributions within the constraints from the systematic uncertainties discussed in section 8.

7.1 Multijet background

Multijet events are produced with a huge cross section via the strong interaction, and therefore give rise to potentially large backgrounds. A first class of MJ background arises from jets or photon conversions misidentified as electrons, or from semileptonic heavy-flavour decays; the 1- and 2-lepton channels are especially sensitive to this class of background. Another class, which affects mostly the 0-lepton channel, arises from large fluctuations in jet energy measurements in the calorimeters, which create “fake” E_T^{miss}. These MJ backgrounds cannot be determined reliably by simulation, and are estimated from data in each of the 0-, 1-, and 2-lepton channels, and in each of the 2- and 3-jet, 0-, 1-, and 2-tag regions.

The MJ background is estimated in the 0-lepton channel using an “ABCD method”, within which the data are divided into four regions based on the min[Δφ(E_T^{miss}, jet)] and Δφ(E_T^{miss}, p_T^{miss}) variables, such that three of the regions are dominated by background. (In the 100–120 GeV p_T^{VH} interval, the likelihood ratio L designed to suppress the MJ background is used instead of min[Δφ(E_T^{miss}, jet)].) For events with real E_T^{miss}, it is expected that the directions of the calorimeter-based and track-based missing transverse momenta, E_T^{miss} and p_T^{miss}, are similar. In events with fake E_T^{miss} arising from a jet energy fluctuation, it is expected that the direction of E_T^{miss} is close to the direction of the poorly measured jet. The signal region (A) is therefore selected with min[Δφ(E_T^{miss}, jet)] > 1.5 and Δφ(E_T^{miss}, p_T^{miss}) < π/2. In region C, the requirement on Δφ(E_T^{miss}, p_T^{miss}) is reversed. In regions B and D, min[Δφ(E_T^{miss}, jet)] < 0.4 is required, with requirements on Δφ(E_T^{miss}, p_T^{miss}) as in regions A and C, respectively. A comparison of the min[Δφ(E_T^{miss}, jet)] distributions for Δφ(E_T^{miss}, p_T^{miss}) above and below π/2 shows that these two variables are only weakly correlated, and this observation is confirmed in a...
multijet event sample simulated with {	extsc{pythia8}}. An MJ template in region A is obtained using events in region C after subtracting the contribution of other backgrounds, taken from simulation. The template is normalised by the ratio of the number of events in region B to that in region D, again after subtracting other backgrounds from those regions. The populations of events in the various regions suffer from low statistical precision after the 2-tag requirement. The $b$-tagging requirement is therefore dropped in regions B, C and D, and an additional $b$-tagging normalisation factor is applied to the resulting template, taken as the fraction of 2-tag events in region D. The MJ background in the signal regions is found to amount to $\sim 1\%$ of the total background.

In the 1-lepton channel, the MJ background is determined separately for the electron and muon sub-channels. For each signal or control region, an MJ-background template is obtained in an MJ-dominated region after subtracting the small remaining contribution from the other backgrounds. The other backgrounds are taken from a simulation improved by scale factors for the various contributions obtained from a preliminary global fit. The MJ-dominated region is obtained by modifying the nominal selection to use medium, instead of tight, leptons and loosening both the track and calorimeter-based isolation criteria. The track-based isolation is changed to the intervals 5\%–12\% and 7\%–50\% for electrons and muons respectively, instead of <4\%; and the calorimeter-based isolation is loosened to <7\% from <4\%. The sample sizes of the MJ-templates are however rather low in the 2-tag regions. Since it is observed that the kinematic properties of the 1-tag and 2-tag events in the MJ-dominated regions are similar, 1-tag events are used to enrich the 2-tag MJ templates. Events in the 1-tag category are promoted to the 2-tag category by assigning to the untagged jet an emulated $MV1c$ value drawn from the appropriate $MV1c$ distribution observed in the corresponding 2-tag MJ template. This distribution depends on the rank (leading or sub-leading) of the untagged jet and on the $MV1c$ value of the tagged jet. To cope with residual differences observed in some distributions between these pseudo-2-tag MJ events and the actual 2-tag MJ events, a reweighting is applied according to the $MV1c$ of the tagged jet and, for the electron sub-channel, according to $\Delta R(jet_1, jet_2)$ and $p_T^{W}$. This procedure is applied in each of the 2- and 3-jet, LL, MM and TT categories. The normalisations of the MJ templates are then obtained from “multijet fits” to the $E_{\text{miss}}^{\text{T}}$ distributions in the 2- and 3-jet, 1- and 2-tag (LL, MM and TT combined) categories, with floating normalisations for the templates of the other background processes. The templates for these other background processes are taken from the improved simulation mentioned above.

The MJ background in the 1-lepton channel is concentrated at low $p_T^{W}$, and in the 2-jet 2-tag sample with $p_T^{W} < 120\text{ GeV}$ it ranges from 11\% of the total background in the LL category to 6\% in the TT category. The main purpose of including the $p_T^{W} < 120\text{ GeV}$ intervals is to provide constraints on the largest backgrounds ($V+\text{jets}$ and $t\bar{t}$) in the global fit. Since the MJ background is twice as large for $p_T^{W} < 120\text{ GeV}$ in the 1-electron sub-channel than in the 1-muon sub-channel, only the 1-muon sub-channel is kept for $p_T^{W} < 120\text{ GeV}$ so as to provide the most reliable constraints on the non-MJ backgrounds. The resulting loss in sensitivity is 0.6\%. For $p_T^{W} > 120\text{ GeV}$, the MJ background is much smaller: 4\% and 2\% in the LL and TT categories, respectively, for 2-jet events.
A template for the MJ background in the 2-electron sub-channel is obtained in a similar way, by loosening identification and isolation requirements. The normalisation is performed by a fit to the dilepton-mass distribution, where the $Z$+jets and MJ components are free parameters, while the other backgrounds (mostly $t\bar{t}$) are taken from the simulation. The MJ normalisation factors are found to be consistent in the 0-, 1- and 2-tag regions. To cope with the reduced size of the 2-tag MJ event sample, a procedure similar to that used in the 1-lepton channel is used, wherein the pretag MJ sample is weighted by its 2-tag fraction and combinations of $MV1c$ values are randomly assigned to the jets according to their distribution in the 2-tag MJ template. In the 2-muon sub-channel, the MJ background is found to be negligible from a comparison between data and MC prediction in the side-bands of the $Z$ mass peak. Altogether, the MJ background amounts to $<1\%$ of the total background in the 2-lepton channel.

7.2 Corrections to the simulation

The large number of events in the 0-tag samples allows for detailed investigations of the modelling of the $V$+jet backgrounds by the version of the SHERPA generator used in this analysis. Given that the search is performed in intervals of $p_T^V$, with the higher intervals providing most of the sensitivity, an accurate modelling of the $p_T^V$ distribution is important. Figure 7(a) shows that the $p_T^W$ spectrum for $W$+jets production in the 1-muon sub-channel is softer in the data than in the simulation. It is found that this mismodelling is strongly correlated with a mismodelling of the $\Delta\phi(jet_1,jet_2)$ distribution, shown in figure 8(a). In order to address this mismodelling, the $Wl$ and $Wcl$ simulations are reweighted based on parameterised fits to the ratio of data to simulation in the $\Delta\phi(jet_1,jet_2)$ variable in the 0-tag region, where these backgrounds dominate. Four separate functions are derived: for the 2- and 3-jet categories and for $p_T^V$ above and below 120 GeV. The reweighted $\Delta\phi(jet_1,jet_2)$ distributions show good agreement between data and simulation (figure 8(b)). This reweighting increases (reduces) by 0.7% (5.6%) the normalisation of the $p_T^W < (>) 120 \text{ GeV}$ region. After this reweighting, the modelling of the whole $p_T^W$ distribution is greatly improved, as can be seen in figure 7(b). This reweighting also improves the modelling of other distributions, most notably the dijet mass. It also improves the modelling in the 1-tag control regions and is therefore applied to the $Wl$ and $Wcl$ backgrounds in all regions of all channels. The numbers of $Wcc$ and $Wb$ background events in the 0- and 1-tag regions are too small to allow conclusive studies of their modelling, so no reweighting is applied to these backgrounds, but an associated systematic uncertainty is assessed instead, as explained in section 8.

A similar, but not identical, procedure is used for the $Z$+jet events in the 2-lepton channel. A $\Delta\phi(jet_1,jet_2)$ reweighting is found to improve the modelling of the $p_T^Z$ distribution in the 0-tag regions. In the signal-depleted 2-tag regions obtained by the exclusion

\footnote{It has indeed been observed that the shape of the $\Delta\phi(jet_1,jet_2)$ distribution in data is better reproduced by NLO generators than by the baseline SHERPA generator used in this analysis [72].}

\footnote{The peak around $\Delta\phi(jet_1,jet_2)=0.7$ comes from the combination of two effects: a rise towards low $\Delta\phi(jet_1,jet_2)$ due to gluon splitting, and a drop towards low $\Delta\phi(jet_1,jet_2)$ due to the two jets becoming unresolved.}
Figure 7. The $p_T^{W}$ distribution observed in data (points with error bars) and expected (histograms) for the 2-jet 0-tag control region of the 1-muon sub-channel (MVA selection), (a) before and (b) after $\Delta\phi(j_{1}, j_{2})$ reweighting. The multijet and simulated-background normalisations are provided by the multijet fits. The size of the statistical uncertainty is indicated by the shaded band. The data-to-background ratio is shown in the lower panel.

Figure 8. The $\Delta\phi(j_{1}, j_{2})$ distribution observed in data (points with error bars) and expected (histograms) for the 2-jet 0-tag control region of the 1-muon sub-channel (MVA selection), (a) before and (b) after reweighting. All $p_T^{W}$ intervals are combined. The multijet and simulated-background normalisations are provided by the multijet fits. The size of the statistical uncertainty is indicated by the shaded band. The data-to-background ratio is shown in the lower panel.
of the 100–150 GeV dijet mass interval, there is no evidence of a need for a $\Delta \phi(\text{jet}_1, \text{jet}_2)$ correction, but the $p_T^Z$ distribution is mismodelled. A dedicated $p_T^Z$ reweighting is therefore determined in the 2-tag regions. Applying the $\Delta \phi(\text{jet}_1, \text{jet}_2)$ reweighting to the $Zl$ component and the $p_T^Z$ reweighting to the $Zc$ and $Zb$ components leads to good modelling also in the 1-tag regions. This procedure is therefore used in all regions of all channels.

It has been observed in an unfolded measurement of the $p_T$ distribution of top quarks from pair production that the POWHEG generator interfaced to PYTHIA predicts too hard a spectrum [73]. A correction accounting for this discrepancy is therefore applied at the level of generated top quarks in the $t\bar{t}$ production process.

7.3 Distributions in the dijet-mass analysis

Distributions of $p_T^V$ and dijet mass are shown in figure 9 and in figures 10 and 11, respectively, for a selection of 2-tag signal regions of the dijet-mass analysis. It can be seen that the background composition in the signal regions varies greatly from channel to channel, with the $p_T^V$ interval, with the jet multiplicity, and with the $b$-tagging category considered. The signal-to-background ratio is larger in the 2-jet and tighter $b$-tagging categories, and lower in the 3-jet and loose $b$-tagging categories.

In the 2-lepton channel, the dominant background is always $Zbb$. There is also a significant contribution from $t\bar{t}$ in the lower $p_T^Z$ intervals, and the relative diboson contribution increases with $p_T^Z$.

For the 1-lepton channel and in the 2-jet samples the combination of $Wbb$ and $t\bar{t}$ accounts for most of the background in the most sensitive MM and TT categories, with the relative contribution of $Wbb$ and dibosons being largest in the tighter $b$-tagging categories and increasing with $p_T^W$. The flavours of the two selected jets from $t\bar{t}$ depend on the reconstructed $p_T^W$ interval. In particular, at high $p_T^W$, when the $b$-quark and the $W$ from a top-quark decay are collimated, there is a large $bc$ contribution, where the $c$-quark comes from the $W \rightarrow cs$ decay. A significant contribution from single-top-quark production processes is also seen. In the 3-jet category, the $t\bar{t}$ contribution is in general dominant, but there are significant contributions from single-top-quark production (mostly in the $Wt$ channel) and from $Wbb$, the latter increasing with $p_T^W$. A non-negligible contribution of MJ background can be seen in the lowest $p_T^W$ intervals of the 2-jet category.

In the 0-lepton channel, the main backgrounds arise from $Zbb$ and $t\bar{t}$, but the $Wbb$ background is also significant. The relative $t\bar{t}$ contribution is largest in the lowest $p_T^V$ intervals, and larger in the 3-jet than in the 2-jet category.

The variations in the background composition between categories allow the global fit to disentangle the rates of the various background sources. The non-negligible contributions from the $Vcl$ and, to a lesser extent, the $Vl$ backgrounds are constrained in the global fit by the LL $b$-tagging categories, and also by the $MV1c$ distributions of the $b$-tagged jet in the 1-tag control regions. The 0-tag control regions are not taken into account in the global fit, but are mainly used to improve the modelling of the $V+$jets backgrounds, as explained in section 7.2.
Figure 9. The $p_T^V$ distribution observed in data (points with error bars) and expected (histograms) for (a) the 2-jet signal regions of the 0-lepton channel with the Medium and Tight $b$-tagging categories (also referred to as MM and TT in the text) combined, (b) the 2-jet signal regions of the 1-lepton channel for the LL category, (c) the 2-jet signal regions of the 1-lepton channel with the MM and TT categories combined, (d) the 3-jet signal regions of the 1-lepton channel with the MM and TT categories combined, and (e) the 2-jet signal regions of the 2-lepton channel with the MM and TT categories combined. The background contributions after the global fit of the dijet-mass analysis are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. Overflow entries are included in the last bin. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 10. The dijet-mass distribution observed in data (points with error bars) and expected (histograms) in the 0-lepton channel with the Medium and Tight b-tagging categories (also referred to as MM and TT in the text) combined for (a) the 2-jet signal region in the $100 < p_T < 120$ GeV interval, (b) the 2-jet signal regions in the three intervals with $p_T > 120$ GeV combined, and (c) the 3-jet signal regions in the three intervals with $p_T > 120$ GeV combined. The background contributions after the global fit of the dijet-mass analysis are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The entries in overflow are included in the last bin. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
**Figure 11.** The dijet-mass distribution observed in data (points with error bars) and expected (histograms) with the Medium and Tight $b$-tagging categories (also referred to as MM and TT in the text) combined and the three intervals with $p_T^{V_T}>120$ GeV combined for (a) the 2-jet signal regions of the 1-lepton channel, (b) the 3-jet signal regions of the 1-lepton channel, (c) the 2-jet signal regions of the 2-lepton channel, and (d) the 3-jet signal regions of the 2-lepton channel. The background contributions after the global fit of the dijet-mass analysis are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The entries in overflow are included in the last bin. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
7.4 Distributions in the multivariate analysis

Distributions of the BDT_{VH} discriminants of the MVA are shown in figures 12 to 14 for 2-tag signal regions in the 2- and 3-jet categories of the 0-, 1- and 2-lepton channels. It can be seen that the backgrounds dominated by light jets and, to a lesser extent, c-jets accumulate at lower values of the BDT_{VH} discriminants, due to the inclusion of the MV1c information as inputs to the BDTs. The composition of the dominant backgrounds accumulating at higher values of the BDT_{VH} discriminant is similar to what was already observed in the 2-tag signal regions of the dijet-mass analysis, namely \( Vbb \) and \( t\bar{t} \), however with a larger contribution of the latter due to the looser requirement on \( \Delta R(\text{jet}_1, \text{jet}_2) \) in the MVA selection.

Distributions of the output of the MV1c \( b \)-tagging algorithm are shown in figure 15 for the \( b \)-tagged jet in the 1-tag control regions of the MVA, in the 2-jet category and for \( p_T^V > 120 \) GeV. In these distributions, the four bins correspond to the four \( b \)-tagging operating points and are ordered from left to right in increasing \( b \)-jet purity. It can be seen that these distributions, which are used in the global fit, provide strong constraints on the \( Vc \) and \( Vl \) backgrounds. As in the dijet-mass analysis, the 0-tag control regions are not used in the global fit.
Figure 12. The $\text{BDT}_{VH}$-discriminant distribution observed in data (points with error bars) and expected (histograms) for the 0-lepton channel combining the 2-tag Medium and Tight $b$-tagging categories (also referred to as MM and TT in the text) for $p_T > 120$ GeV for (a) 2-jet events and (b) 3-jet events. The background contributions after the global fit of the MVA are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 13. The BDT_{VH}-discriminant distribution observed in data (points with error bars) and expected (histograms) for the 2-tag signal regions of the 1-lepton channel for (a) 2-jet events with the Medium and Tight b-tagging categories (also referred to as MM and TT in the text) combined and with $p_T^{W} \leq 120$ GeV, (b) MM 2-jet events with $p_T^{W} > 120$ GeV, (c) TT 2-jet events with $p_T^{W} > 120$ GeV, and (d) MM and TT combined 3-jet events with $p_T^{W} > 120$ GeV. The background contributions after the global fit of the MVA are shown as filled histograms. The Higgs boson signal ($m_{H} = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 14. The BDT $V^{H}$-discriminant distribution observed in data (points with error bars) and expected (histograms) for the 2-lepton channel combining the 2-tag Medium and Tight $b$-tagging categories (also referred to as MM and TT in the text) for (a) 2-jet events with $p_{T}^{Z} \leq 120$ GeV, (b) 2-jet events with $p_{T}^{Z} > 120$ GeV, and (c) 3-jet events with $p_{T}^{Z} > 120$ GeV. The background contributions after the global fit of the MVA are shown as filled histograms. The Higgs boson signal ($m_{H} = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
Figure 15. Distribution of the output of the MV1c b-tagging algorithm observed in data (points with error bars) and expected (histograms) for the 1-tag control regions of the MVA in the 2-jet category with $p_T > 120$ GeV (a) in the 0-lepton channel, (b) in the 1-lepton channel, and (c) in the 2-lepton channel. The left bin boundaries denote the operating points (MV1c(b) OP) as defined in section 4, corresponding to b-tagging efficiencies of 80%, 70%, 60%, 50%, i.e., the b-jet purity increases from left to right. The background contributions after the global fit of the MVA are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $µ = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
8 Systematic uncertainties

The systematic uncertainties discussed in this section are: those of experimental origin; those related to the multijet background estimation; and those associated with the modelling of the simulated backgrounds and Higgs boson signal.

8.1 Experimental uncertainties

All relevant experimental systematic uncertainties are considered, such as those affecting the trigger selection, the object reconstruction and identification, and the object energy and momentum calibrations and resolutions. The most relevant ones are discussed in the following.

For the $E_T^{\text{miss}}$ trigger, an efficiency correction is derived from $W \rightarrow \mu\nu+\text{jets}$ and $Z \rightarrow \mu^+\mu^-+\text{jets}$ events. This correction amounts to 4.5% for events with an $E_T^{\text{miss}}$ of 100 GeV, the threshold required in the analysis, and is below 1% for $E_T^{\text{miss}} > 120$ GeV. The associated uncertainties arise from the statistical uncertainties of this method and differences observed in the two event classes. They are very small (below 1%) for the high $E_T^{\text{miss}}$ (and thus high $p_T^V$) intervals, and reach about 3% for the low $E_T^{\text{miss}}$ interval of the 0-lepton channel ($100–120$ GeV).

For electrons and muons, uncertainties associated with the corrections for the trigger, reconstruction, identification and isolation efficiencies are taken into account. Uncertainties on energy and resolution corrections of the leptons are also considered. The impact of these uncertainties is very small, typically less than 1%.

Several sources contribute to the uncertainty of the jet energy scale (JES) \cite{61} related e.g. to uncertainties from in situ calibration analyses, pile-up-dependent corrections and the flavour composition of jets in different event classes. After being decomposed into uncorrelated components, these are treated as independent sources in the analysis. The total relative systematic uncertainties on the JES range from about 3% to 1% for central jets with a $p_T$ of 20 GeV and 1 TeV, respectively. An additional specific uncertainty of about 1%–2% affects the energy calibration of $b$-jets. Small uncertainties on the corrections applied to improve the dijet-mass resolution are also included. Corrections and uncertainties are also considered for the jet energy resolution (JER) \cite{74}, with a separate contribution for $b$-jets. The total relative systematic uncertainty on the JER ranges from about 10% to 20%, depending on the $\eta$ range, for jets with $p_T = 20$ GeV to less than 5% for jets with $p_T > 200$ GeV.

The JES uncertainties are propagated to the $E_T^{\text{miss}}$, as are the much smaller uncertainties related to the energy and momentum calibration of leptons. An uncertainty on the $E_T^{\text{miss}}$ also comes from the uncertainties on the energy calibration (8%) and resolution (2.5%) of calorimeter energy clusters not associated with any reconstructed object \cite{66}.

The $b$-tagging efficiencies for the different jet flavours are measured in both data and simulation using dedicated event samples \cite{63, 64}. The $b$-tagging efficiencies for simulated jets are corrected within intervals between operating points by MC-to-data SFs, which depend on the jet kinematics. For $b$-jets, the precision is driven by an analysis of $t\bar{t}$ events in final states containing two leptons. The MC-to-data SFs are close to unity,
with uncertainties at the level of 2–3\% over most of the jet $p_T$ range, reaching 5\% for $p_T = 20\text{ GeV}$ and 8\% above 200 GeV. The uncertainties, which depend on $p_T$ and on the interval between operating points, are decomposed into uncorrelated components and the ten most significant ones are kept in the analysis. It was checked that the neglected components have a negligible impact. The uncertainties on $c$-jets are decomposed into 15 components, and the uncertainties on light jets, to which the analysis is much less sensitive, are decomposed into ten components, accounted for in $p_T$ and $\eta$ ranges. For $b$- and $c$-jets further uncertainties are added for the application of the additional MC-to-MC SFs to obtain generator specific MC-to-data SFs as explained in section 4. Half of the correction is used as systematic uncertainty. As discussed in section 4, a correction to $c$-jets in the $Vcc$ samples, for which parameterised tagging is used, is applied at low $\Delta R$ to the closest jet. Half of this correction is assigned as a systematic uncertainty.

The uncertainty on the integrated luminosity is 2.8\%. It is derived, following the same methodology as that described in ref. [20], from a preliminary calibration of the luminosity scale derived from beam-separation scans performed in November 2012. It is applied to the signal and backgrounds estimates that are taken from simulation. A 4\% uncertainty on the average number of interactions per bunch crossing is taken into account.

8.2 Uncertainties on the multijet backgrounds

In the 0-lepton channel, the robustness of the MJ background estimation is assessed by varying the $\min[\Delta \phi(E_T^{\text{miss}}, \text{jet})]$ values defining the B and D regions of the ABCD method, and by replacing the $b$-tagging fractions measured in region D by those measured in region B. A systematic uncertainty of 100\% is assessed for this small (∼1\%) background, uncorrelated between 2- and 3-jet, 1- and 2-$b$-tag categories. The MJ background in the 2-lepton channel is also at the per-cent level, and an uncertainty of 100\% is assigned.

In the 1-lepton channel, normalisation uncertainties arise from the statistical uncertainties of the multijet fits and from uncertainties on the non-MJ background subtractions performed to construct the MJ templates. Normalisation uncertainties are also assessed in the LL, MM and TT categories to cover differences between multijet fits performed inclusively in the 2-tag regions and in the individual categories. In the 2-jet 2-tag region of the electron sub-channel, the overall normalisation uncertainties amount to 11\%, 14\% and 22\% in the LL, MM and TT categories, respectively. In the muon sub-channel, the corresponding uncertainties are about three times larger because of the smaller size of the MJ-enriched samples.

In the 1-lepton channel, shape uncertainties are assessed in the various regions by comparison of evaluations obtained using MJ-enriched samples defined by isolation requirements different from those applied in the nominal selections. In the electron sub-channel, an alternative template is constructed with a track-based isolation in the 12\% to 50\% interval, and another alternative template with a calorimeter-based isolation in the 0\% to 4\% interval. In the muon sub-channel, the results obtained with the nominal MJ template are compared with those obtained with tighter or looser isolation requirements, defined by track-based isolation intervals of 7\%–9.5\% and 9.5\%–50\%, respectively. Furthermore,
half of the $\Delta R(jet_1, jet_2)$ and $p_T^{V}$ reweightings mentioned in section 7.1 for the electron sub-channel are taken as systematic uncertainties.

8.3 Uncertainties on the modelling of the simulated backgrounds

The physics-modelling systematic uncertainties evaluated focus on the quantities that are used in the global fit, i.e., those affecting the jet multiplicities, the $p_T^{V}$ distributions, the flavour composition and the $m_{bb}$ distributions. For the MVA, systematic uncertainties affecting the other variables used as inputs to the BDTs are also considered. Whenever possible, dedicated control regions are used to extract information directly from the data. This is the case for $Z+jets$ and $W+light$ jets. In other cases, uncertainties are assessed by comparison of MC predictions based on a variety of generators with the nominal ones.

Details of the assessment of systematic uncertainties are provided below in the context of the MVA. When systematic uncertainties are derived from a comparison between generators, all relevant variables are considered independently. The variable showing the largest discrepancy in some generator with respect to the nominal generator is assigned an uncertainty covering this discrepancy, which is symmetrised. If, once propagated to the BDT$_{VH}$ discriminant, this uncertainty is sufficient to cover all variations observed with the different generators, it is considered to be sufficient. If not, an uncertainty is considered in addition on the next most discrepant variable and the procedure is iterated until all variations of the BDT$_{VH}$ discriminant are covered by the assigned uncertainties.

A given source of systematic uncertainty can affect different analysis regions. Whether such an uncertainty should be treated as correlated or not depends on whether constraints resulting from the global fit should be propagated from one region to another. Details of the procedures leading to such decisions are provided in section 9.2.

A summary of the systematic uncertainties affecting the modelling of the backgrounds can be found in table 5.

Top-quark-pair background. As explained in section 7, the top-quark $p_T$ distribution is reweighted at generator level to bring it into agreement with measurement [73]. A systematic uncertainty amounting to half of this correction is assigned, correlated across channels.

The predictions of the nominal $t\bar{t}$ generator (POWHEG+PYTHIA) are compared, focussing on the 1-lepton channel selection, with those obtained using a variety of generators differing by the PDF choice (POWHEG+PYTHIA with HERAPDF [75]), by the parton showering and hadronisation scheme (POWHEG+HERWIG), by the implementation of the NLO matrix element and the matching scheme (MC@NLO [76]+HERWIG), by the amount of initial- and final-state radiation (ISR/FSR) using AcerMC+PYTHIA, or by the implementation of higher-order tree-level matrix elements (ALPGEN [77]+PYTHIA). It is found that, in general, the largest deviations are observed for ALPGEN, which is therefore used to assess further systematic uncertainties as explained below.

In the global fit, the normalisation of the $t\bar{t}$ background in the 2-jet category is left floating freely, independently in each of the lepton channels. An uncertainty of 20% on the 3-to-2-jet ratio is estimated from the generator comparisons explained above. In the
global fit, this uncertainty is treated as correlated between the 0- and 1-lepton channels, and uncorrelated with the 2-lepton channel.

The shape of the $m_{bb}$ distribution is also studied with the same set of generators, leading to correlated shape uncertainties for 2- and 3-jet events, and for $p_T^V < 120$ GeV and $p_T^V > 120$ GeV. The associated variation is larger in the higher $p_T^V$ interval: for 2-jet events, when it increases the distribution by 3% for $m_{bb} = 50$ GeV, it decreases it by 1% at 200 GeV; the effect is similar, but of opposite sign, for 3-jet events.

The same procedure is used for the $p_T^V$ distribution, from which a 7.5% uncertainty is assessed on the normalisation of the $p_T^V > 120$ GeV interval. Finally, the same approach calls for a shape systematic uncertainty on the $E_T^{\text{miss}}$ distribution in the 1-lepton channel, different but correlated between $p_T^V < 120$ GeV and $p_T^V > 120$ GeV. This uncertainty is not applied in the 0- and 2-lepton channels.

**Single-top-quark background.** The theoretical uncertainties on the cross sections of the three processes contributing to single-top production are 4%, 4%, and 7% for the s-channel, t-channel, and Wt production, respectively [78].

The predictions of the nominal generators (POWHEG+PYTHIA for the s-channel and for Wt production; ACERMC+PYTHIA for the t-channel) are compared, after the 1-lepton channel selection, with those obtained using a variety of generators. For the s-channel, the comparison is made with ACERMC and MC@NLO; for Wt production with ACERMC, POWHEG+HERWIG, and MC@NLO; and for the t-channel with aMC@NLO [81, 82] + HERWIG. For all three processes, the impact of ISR/FSR is evaluated using ACERMC. For Wt production, there are interference effects with $tt$ production, which need to be considered. Two methods are available for this: the Diagram Removal (DR) and the Diagram Subtraction (DS) schemes [83]. The former is used in the nominal generation, and the second for comparison.

Uncertainties on the acceptance for each of the three processes are taken as the largest deviations observed, separately for $p_T^V < 120$ GeV and $p_T^V > 120$ GeV, and for 2- and 3-jet events. They can be as large as 52% for 2-jet events in the t-channel at low $p_T^V$, of the order of 5% for Wt production (except for 3-jet events at high $p_T^V$: 15%), and typically 20% for the s-channel.

In addition to the acceptance uncertainties, the effects of the model variations described above on variables input to the BDT are evaluated and three shape systematic uncertainties are found to be needed in Wt production. The first uncertainty is on the shape of the $m_{bb}$ distribution in the high $p_T^V$ interval for 2-jet events where, when a shift from the nominal model increases the rate by 20% for $m_{bb} = 50$ GeV, it decreases it by 40% at 200 GeV. A second uncertainty is on the $m_{bb}$ shape for 3-jet events, where the corresponding shifts are 25% and 20%. Finally, a third uncertainty is on the $p_T$ distribution of the leading jet in the low $p_T^V$ interval for 2-jet events.

---

Event generation with aMC@NLO is based on the MC@NLO formalism and the MADGRAPH-5 framework [79, 80].
**Z+jets background.** As explained in section 7, \( \Delta\phi(jet_1, jet_2) \) and \( p_T^Z \) reweightings are applied to the \( Zl \) and \( Zc + Zb \) components, respectively. For the \( \Delta\phi(jet_1, jet_2) \) reweighting, a systematic uncertainty amounting to half of the correction is assigned to the \( Zl \) component, while an uncertainty amounting to the full correction is assigned to the \( Zc + Zb \) components. This is done separately for 2- and 3-jet events, and all these uncertainties are treated as uncorrelated. For the \( p_T^Z \) reweighting, uncorrelated systematic uncertainties of half the correction are assigned to the \( Zl \) component, while an uncertainty amounting to the full correction is assigned to the \( Zc + Zb \) components. The notation \( Zc + Zb \) is meant to indicate that a systematic uncertainty is treated as correlated between the \( Zc \) and \( Zb \) components.

The normalisation and the 3-to-2-jet ratio for the \( Zl \) background are determined from data in the 0-tag region of the 2-lepton channel, both with an uncertainty of 5%. The normalisations of the \( Zcl \) and \( Zbb \) backgrounds are left free in the global fit. The uncertainties on the 3-to-2-jet ratios for the \( Zcl \) and \( Z+hf \) components are assessed through a comparison of ALPGEN with the nominal SHERPA generator in the 2-tag region of the 2-lepton channel; these are 26% for \( Zcl \) and 20% for \( Z+hf \). The same procedure is used to estimate uncertainties on the flavour fractions within \( Z+hf \) events, yielding 12% for each of \( bl/bb, cc/bb \) and \( bc/bb \), with \( bl/bb \) uncorrelated between 2- and 3-jet samples.

The shape of the \( m_{bb} \) distribution is compared between data and simulation in the 2-tag region of the 2-lepton channel, excluding the 100–150 GeV range, from which a shape uncertainty is derived that, when it increases the dijet-mass distribution by 3% at 50 GeV, it decreases it by 5% at 200 GeV. This uncertainty is applied uncorrelated to the \( Zl \) and \( Zb + Zc \) components. The differences between ALPGEN and SHERPA are covered by this uncertainty.

**W+jets background.** As explained in section 7, a \( \Delta\phi(jet_1, jet_2) \) reweighting is applied to the \( Wl \) and \( Wcl \) components. Uncorrelated systematic uncertainties amounting to half of the correction are assigned to these two components, for each of the 2- and 3-jet categories. For the \( Wcc + Wb \) component, no reweighting is applied but a systematic uncertainty is assigned, equal to the full correction applied to the \( Wl \) and \( Wcl \) components, uncorrelated between 2- and 3-jet events.

The normalisation and the 3-to-2-jet ratio for the \( Wl \) background are taken directly from simulation, both with a 10% uncertainty. This is based on the agreement observed between data and prediction in the 0-tag sample. The 3-to-2-jet ratio for the \( Wcl \) background is also assigned an uncertainty of 10%. The normalisations of the \( Wcl \) and \( Wbb \) backgrounds are left free in the global fit.

To assign further uncertainties on the \( Wbb \) background, for which dedicated control regions are not available in the data, extensive comparisons are performed at generator level, with kinematic selections mimicking those applied after reconstruction. The predictions of the SHERPA generator are compared to those of POWHEG+PYTHIA8, of aMC@NLO+HERWIG++ [84] and of ALPGEN+HERWIG. Comparisons are also made between samples generated with aMC@NLO with renormalisation (\( \mu_R \)) and factorisation (\( \mu_F \)) scales\(^8\) independently modified by factors of 2 or 0.5 and also with different PDF sets

\(^8\)The nominal scales are taken as \( \mu_R = \mu_F = [m_W^2 + p_T(W)^2 + m_b^2 + (p_T(b))^2 + p_T(\bar{b})^2)/2]^{1/2} \).
As a result, a 10% uncertainty is assigned to the 3-to-2-jet ratio, taken as correlated between all $W+hf$ processes. Shape uncertainties are also assessed for the $m_{bb}$ and $p_T^W$ distributions. When the former increases the dijet-mass distribution by 23% at 50 GeV, it decreases it by 28% at 200 GeV. It is taken as uncorrelated for $Wl$, $Wcl$, $Wbb + Wcc$ and $Wbl + Wbc$. For $Wbb + Wcc$, it is furthermore uncorrelated among $p_T^W$ intervals (with the three highest intervals correlated for the dijet-mass analysis). When the latter shape uncertainty increases the $p_T^W$ distribution by 9% for $p_T^W = 50$ GeV, it decreases it by 23% at 200 GeV. It is taken as correlated for all $W+hf$ processes, and uncorrelated between the 2- and 3-jet samples.

Predictions using the inclusive production of all flavours by sherpa and alpgen are compared after full reconstruction and event selection to assign uncertainties on the flavour fractions that take properly into account heavy-flavour production at both the matrix-element and parton-shower levels. (For alpgen, the production of light flavours and heavy flavours are performed separately at the matrix-element level; a dedicated procedure, based on the $\Delta R$ separation between $b$-partons, is used to remove the overlap between $b\bar{b}$ pairs produced at the matrix-element and parton-shower levels.) The following uncertainties are assigned in the $W+hf$ samples: 35% for $bl/bb$ and 12% for each of $bc/bb$ and $cc/bb$. The uncertainty on $bl/bb$ is uncorrelated between $p_T$ intervals (with the three highest intervals correlated for the dijet-mass analysis).

**Diboson background.** The uncertainties on the cross sections for diboson production ($WW$, $WZ$ and $ZZ$) are assessed at parton level using mcfm at NLO in QCD. The sources of uncertainty considered are the renormalisation and factorisation scales and the choice of PDFs. The nominal scales are dynamically set to half of the invariant mass of the diboson system and the nominal PDFs are the CT10 set.

The scale uncertainties are evaluated by varying simultaneously $\mu_R$ and $\mu_F$ by factors of 2 or 0.5. Since the analysis is performed in $p_T^V$ intervals and in exclusive 2- and 3-jet categories, the uncertainties are evaluated for each channel separately in those intervals and categories (2 and 3 final-state partons within the nominal selection acceptance) following the prescription of ref. [86]. This procedure leads, in each $p_T^V$ interval, to two uncorrelated uncertainties in the 2-jet category, one for 2+3 jets inclusively and one associated with the removal of 3-jet events, and to one in the 3-jet category anti-correlated with the latter uncertainty in the 2-jet category. These uncertainties are largest at high $p_T^V$. For $p_T^V > 200$ GeV, the two uncertainties affecting the 2-jet category can be as large as 29% and 22% in the $WZ$ channel, roughly half this size in the $ZZ$ channel and intermediate for $WW$; and the uncertainty affecting the 3-jet category is about 17% in all channels.

The uncertainties due to the PDF choice are evaluated according to the PDF4LHC recommendation [87], i.e., using the envelope of predictions from the CT10, MSTW2008NLO, and NNPDF2.3 PDF sets and their associated uncertainties. They range from 2% to 4%, with no $p_T^V$ dependence observed.

The shape of the reconstructed $Z \rightarrow b\bar{b}$ lineshape in $VZ$ production is affected by the parton-shower and hadronisation model. A shape-only systematic uncertainty is assessed by comparing the lineshapes obtained with the nominal powheg+pythia8 gener-
ator and with HERWIG. The relative difference between the shapes is 20% for a dijet mass around 125 GeV.

8.4 Uncertainties on the signal modelling

The $q\bar{q} \rightarrow WH$, $q\bar{q} \rightarrow ZH$, and $gg \rightarrow ZH$ signal samples are normalised respectively to their inclusive cross sections as explained in section 3. The uncertainties on these cross sections [88] include those arising from the choice of scales $\mu_R$ and $\mu_F$ and of PDFs.

The scale uncertainty is 1% for $WH$ production. It is larger (3%) for $ZH$ production, due to the contribution of the gluon-gluon initiated process. Under the assumption that the scale uncertainties are similar (1%) for $qq \rightarrow WH$ and $qq \rightarrow ZH$, a conservative uncertainty of 50% is inferred for $gg \rightarrow ZH$. The same procedure leads to PDF uncertainties of 2.4% for $qq \rightarrow (W/Z)H$ and 17% for $gg \rightarrow ZH$. The relative uncertainty on the Higgs boson branching ratio to $bb$ is 3.3% for $m_H = 125$ GeV [11]. The contribution of decays to final states other than $bb$ is verified to amount to less than 1% after selection.

Acceptance uncertainties due to the choice of scales are determined from signal samples generated with POWHEG interfaced to PYTHIA8, with $\mu_R$ and $\mu_F$ varied independently by factors of 2 or 0.5. The procedure advocated in ref. [86] is used, after kinematic selections applied at generator level, leading to acceptance uncertainties of 3.0%, 3.4% and 1.5% for $qq \rightarrow WH$, $qq \rightarrow ZH$ and $gg \rightarrow ZH$, respectively, for the 2- and 3-jet categories combined, and of 4.2%, 3.6% and 3.3% for the 3-jet category. The latter uncertainty is anti-correlated with an acceptance uncertainty associated with the removal of 3-jet events from the 2+3-jet category to form the 2-jet category. In addition, the $p_T^V$ spectrum is seen to be affected, and shape uncertainties are derived. For the $qq \rightarrow (W/Z)H$ samples, when they increase the distribution by 1% for $p_T^V = 50$ GeV, they decrease it by 3% at 200 GeV. These variations are 2% and 8%, respectively, for the $gg \rightarrow ZH$ samples.

Acceptance uncertainties due to the PDF choice are determined in a similar way, following the PDF4LHC prescription. They range from 2% in the 2-jet $gg \rightarrow ZH$ samples to 5% in the 3-jet $qq \rightarrow ZH$ samples. There is no evidence of a need for $p_T^V$ shape uncertainties related to the PDFs.

The applied uncertainties on the shape of the $p_T^V$ spectrum associated with the NLO electroweak corrections [38] are typically at the level of 2%, increasing with $p_T^V$ to reach 2.5% in the highest $p_T^V$ interval.

The effect of the underlying-event modelling is found to be negligible, using various PYTHIA tunes. The effect of the parton-shower modelling is examined by comparison of simulations by POWHEG interfaced with PYTHIA8 and with HERWIG. Acceptance variations of 8% are seen, except for 3-jet events in the $p_T^V > 120$ GeV interval, where the variation is at the level of 13%. These variations are taken as systematic uncertainties.

A summary of the systematic uncertainties affecting the modelling of the Higgs boson signal is given in table 5.
<table>
<thead>
<tr>
<th>Signal</th>
<th>1% ((q\bar{q})), 50% ((gg))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross section (scale)</td>
<td>2.4% ((q\bar{q})), 17% ((gg))</td>
</tr>
<tr>
<td>Cross section (PDF)</td>
<td></td>
</tr>
<tr>
<td>Branching ratio</td>
<td>3.3%</td>
</tr>
<tr>
<td>Acceptance (scale)</td>
<td>1.5%–3.3%</td>
</tr>
<tr>
<td>3-jet acceptance (scale)</td>
<td></td>
</tr>
<tr>
<td>(p_T^V) shape (scale)</td>
<td>3.3%–4.2%</td>
</tr>
<tr>
<td>Acceptance (PDF)</td>
<td></td>
</tr>
<tr>
<td>(p_T^V) shape (NLO EW correction)</td>
<td></td>
</tr>
<tr>
<td>Acceptance (parton shower)</td>
<td>8%–13%</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>(Z+\text{jets})</td>
<td></td>
</tr>
<tr>
<td>(Zl) normalisation, 3/2-jet ratio</td>
<td>5%</td>
</tr>
<tr>
<td>(Zcl) 3/2-jet ratio</td>
<td>26%</td>
</tr>
<tr>
<td>(Z+hf) 3/2-jet ratio</td>
<td>20%</td>
</tr>
<tr>
<td>(Z+hf/Zbb) ratio</td>
<td>12%</td>
</tr>
<tr>
<td>(\Delta \phi(\text{jet}_1, \text{jet}<em>2), p_T^V, m</em>{bb})</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>(W+\text{jets})</td>
<td></td>
</tr>
<tr>
<td>(Wl) normalisation, 3/2-jet ratio</td>
<td>10%</td>
</tr>
<tr>
<td>(Wcl, W+hf) 3/2-jet ratio</td>
<td>10%</td>
</tr>
<tr>
<td>(Wbl/Wbb) ratio</td>
<td>35%</td>
</tr>
<tr>
<td>(Wbc/Wbb, Wcc/Wbb) ratio</td>
<td>12%</td>
</tr>
<tr>
<td>(\Delta \phi(\text{jet}_1, \text{jet}<em>2), p_T^V, m</em>{bb})</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>(t\bar{t})</td>
<td></td>
</tr>
<tr>
<td>3/2-jet ratio</td>
<td>20%</td>
</tr>
<tr>
<td>High/low-(p_T^V) ratio</td>
<td>7.5%</td>
</tr>
<tr>
<td>Top-quark (p_T, m_{bb}, E_T^{\text{miss}})</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Single top</td>
<td></td>
</tr>
<tr>
<td>Cross section (generator)</td>
<td>4% ((s-, t-\text{channel})), 7% ((Wt))</td>
</tr>
<tr>
<td>Acceptance (generator)</td>
<td>3%–52%</td>
</tr>
<tr>
<td>(m_{bb}, p_T^{b_1})</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Diboson</td>
<td></td>
</tr>
<tr>
<td>Cross section and acceptance (scale)</td>
<td>3%–29%</td>
</tr>
<tr>
<td>Cross section and acceptance (PDF)</td>
<td>2%–4%</td>
</tr>
<tr>
<td>(m_{bb})</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Multijet</td>
<td></td>
</tr>
<tr>
<td>0-, 2-lepton channels normalisation</td>
<td>100%</td>
</tr>
<tr>
<td>1-lepton channel normalisation</td>
<td>2%–60%</td>
</tr>
<tr>
<td>Template variations, reweighting</td>
<td>S</td>
</tr>
</tbody>
</table>

**Table 5.** Summary of the systematic uncertainties on the signal and background modelling. An “S” symbol is used when only a shape uncertainty is assessed.
9 Statistical procedure

9.1 General aspects

A statistical fitting procedure based on the Roostats framework [89, 90] is used to extract the signal strength from the data. The signal strength is a parameter, \( \mu \), that multiplies the SM Higgs boson production cross section times branching ratio into \( b\bar{b} \). A binned likelihood function is constructed as the product of Poisson-probability terms over the bins of the input distributions involving the numbers of data events and the expected signal and background yields, taking into account the effects of the floating background normalisations and the systematic uncertainties.

The different regions entering the likelihood fit are summarised in table 6. In the dijet-mass analysis, the inputs to the “global fit” are the \( m_{bb} \) distributions in the 81 2-tag signal regions defined by three channels (0, 1 or 2 leptons), up to five \( p_{T}^{V} \) intervals, two number-of-jet categories (2 or 3), and three \( b \)-tagging categories (LL, MM and TT). Here and in the rest of this section, \( m_{bb} \) distributions are to be understood as transformed distributions, as explained in section 6. In the MVA, the inputs are the \( BDT_{VH} \) discriminants in the 24 2-tag signal regions defined by the three lepton channels, up to two \( p_{T}^{V} \) intervals, the two number-of-jet categories, and \( b \)-tagging categories. In the 1-lepton channel, the \( b \)-tagging categories are LL, MM and TT. In the 0- and 2-lepton channels, they are the LL category and a combined MM and TT category (MM+TT).\(^9\) These \( BDT_{VH} \)-discriminant distributions are supplemented by the three \( m_{bb} \) distributions in the 100–120 GeV \( p_{T}^{V} \) interval of the 2-jet 2-tag categories (LL, MM, and TT) of the 0-lepton channel. For the MVA, additional inputs are the \( MV_{1c} \) distributions of the \( b \)-tagged jet in the 11 1-tag control regions of the MVA selection and in the 100–120 GeV \( p_{T}^{V} \) interval of the 2-jet 1-tag category of the 0-lepton channel. In the dijet-mass analysis, the \( MV_{1c} \) distributions are combined in each of the \( p_{T}^{V} < 120 \) GeV and \( p_{T}^{V} > 120 \) GeV intervals, which also results in 11 1-tag control regions. Altogether, there are 584 \( m_{bb} \) and \( MV_{1c} \) bins in the 92 regions of the dijet-mass analysis, and 251 \( BDT_{VH} \)-discriminant and \( MV_{1c} \) bins in the 38 regions of the MVA, to be used in the global fits.

The impact of systematic uncertainties on the signal and background expectations is described by nuisance parameters (NPs), \( \theta \), which are constrained by Gaussian or log-normal probability density functions, the latter being used for normalisation uncertainties to prevent normalisation factors from becoming negative in the fit. The expected numbers of signal and background events in each bin are functions of \( \theta \). The parameterisation of each NP is chosen such that the predicted signal and background yields in each bin are log-normally distributed for a normally distributed \( \theta \). For each NP, the prior is added as a penalty term to the likelihood, \( L(\mu, \theta) \), which decreases it as soon as \( \theta \) is shifted away from its nominal value. The statistical uncertainties of background predictions from simulation are included through bin-by-bin nuisance parameters.

\(^9\)While keeping distinct MM and TT categories in the 1-lepton channel improves the sensitivity, this is not observed for the 0- and 2-lepton channels. Keeping the LL category separated from the others improves the sensitivity in all lepton channels.
Table 6. The distributions used in each region by the likelihood fit in the dijet-mass analysis and in the MVA applied to the 8 TeV data. Here, “BDT” stands for “BDT_{VH} discriminant”. For each entry listed, there are additional divisions into $p_T^V$ intervals: five in the dijet-mass analysis and two in the MVA, as shown in table 2. These distributions are input to the fit for the 2-jet and 3-jet categories separately, except in the low $p_T^V$ interval (100–120 GeV) of the 0-lepton channel where only the 2-jet category is used. In the 0- and 2-lepton channels, the MM and TT 2-tag categories are combined in the MVA. (* In the low $p_T^V$ interval of the 0-lepton channel, the MVA uses the $m_{bb}$ distributions in the LL, MM and TT 2-tag categories as well as the $MV_1c$ distribution in the 1-tag category.

The test statistic $q_\mu$ is then constructed from the profile likelihood ratio

$$ q_\mu = -2 \ln \Lambda_\mu \quad \text{with} \quad \Lambda_\mu = \mathcal{L}(\mu, \hat{\theta})/\mathcal{L}(\hat{\mu}, \hat{\theta}), $$

where $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximise the likelihood with the constraint $0 \leq \hat{\mu} \leq \mu$, and $\hat{\theta}$ are the nuisance parameter values that maximise the likelihood for a given $\mu$. This test statistic is used for exclusion intervals derived with the CLs method \[91, 92\]. To measure the compatibility of the background-only hypothesis with the observed data, the test statistic used is $q_0 = -2 \ln \Lambda_0$. The results are presented in terms of: the 95% confidence level (CL) upper limit on the signal strength; the probability $p_0$ of the background-only hypothesis; and the best-fit signal-strength value $\hat{\mu}$ with its associated uncertainty $\sigma_\mu$. The fitted $\hat{\mu}$ value is obtained by maximising the likelihood function with respect to all parameters. The uncertainty $\sigma_\mu$ is obtained from the variation of $2 \ln \Lambda_\mu$ by one unit, where $\Lambda_\mu$ is now defined without the constraint $0 \leq \hat{\mu} \leq \mu$. Expected results are obtained in the same way as the observed results by replacing the data in each input bin by the expectation from simulation with all NPs set to their best-fit values, as obtained from the fit to the data.\[10\]

While the analysis is optimised for a Higgs boson of mass 125 GeV, results are also extracted for other masses. These are obtained without any change to the dijet-mass analysis, except for the binning of the transformed $m_{bb}$ distribution, which is reoptimised. For the MVA, it is observed that the performance degrades for masses away from 125 GeV, for which the BDTs are trained. This is largely due to the fact that $m_{bb}$ is an input to the BDTs. The MVA results for other masses are therefore obtained using BDTs retrained for each of the masses tested at 5 GeV intervals between 100 and 150 GeV. The details provided in the rest of this section refer to the analysis performed for a Higgs boson mass of 125 GeV.

\[10\]This type of pseudo-data sample is referred to as an Asimov dataset in ref. \[92\].
<table>
<thead>
<tr>
<th>Process</th>
<th>Scale factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$ 0-lepton</td>
<td>1.36 ± 0.14</td>
</tr>
<tr>
<td>$t\bar{t}$ 1-lepton</td>
<td>1.12 ± 0.09</td>
</tr>
<tr>
<td>$t\bar{t}$ 2-lepton</td>
<td>0.99 ± 0.04</td>
</tr>
<tr>
<td>$Wbb$</td>
<td>0.83 ± 0.15</td>
</tr>
<tr>
<td>$Wcl$</td>
<td>1.14 ± 0.10</td>
</tr>
<tr>
<td>$Zbb$</td>
<td>1.09 ± 0.05</td>
</tr>
<tr>
<td>$Zcl$</td>
<td>0.88 ± 0.12</td>
</tr>
</tbody>
</table>

Table 7. Factors applied to the nominal normalisations of the $t\bar{t}$, $Wbb$, $Wcl$, $Zbb$, and $Zcl$ backgrounds, as obtained from the global MVA fit to the 8 TeV data. The $t\bar{t}$ background is normalised in the 2-jet category independently in each of the lepton channels. The errors include the statistical and systematic uncertainties.

9.2 Technical details

The data have sufficient statistical power to constrain the largest background-normalisation NPs, which are left free to float in the fit. This applies to the $t\bar{t}$, $Wbb$, $Wcl$, $Zbb$, and $Zcl$ processes. The corresponding factors applied to the nominal background normalisations as resulting from the global fit of the MVA to the 8 TeV data, are shown in table 7. As stated in section 8, the $t\bar{t}$ background is normalised in the 2-jet category independently in each of the lepton channels. The reason for uncorrelating the normalisations in the three lepton channels is that the regions of phase space probed in the 2-jet category are very different between the three channels. In the 2-lepton channel, the $t\bar{t}$ background is almost entirely due to events in which both top quarks decay into $(W \rightarrow \ell \nu)b$ (fully leptonic decays) with all final-state objects detected (apart from the neutrinos). In the 1-lepton channel, it is in part due to fully leptonic decays with one of the leptons (often a $\tau$ lepton) undetected, and in part to cases where one of the top quarks decays as above and the other into $(W \rightarrow q\bar{q}')b$ (semileptonic decays) with a missed light-quark jet. Finally, in the 0-lepton channel, the main contributions are from fully leptonic decays with the two leptons undetected and from semileptonic decays with a missed lepton and a missed light-quark jet; here again, the missed leptons are often $\tau$ leptons. Furthermore, the $p_T^V$ range probed is different in the 0-lepton channel: $p_T^V > 100$ GeV in contrast to being inclusive in the 1- and 2-lepton channels.

As described in detail in section 8, a large number of sources of systematic uncertainty are considered. The number of nuisance parameters is even larger because care is taken to appropriately uncorrelate the impact of the same source of systematic uncertainty across background processes or across regions accessing very different parts of phase space. This avoids unduly propagating constraints. For instance, the $t\bar{t}$ background contributes quite differently in the 2-tag 3-jet regions of the 0- and 1-lepton channels on one side, and of the 2-lepton channel on the other. In the 0- and 1-lepton channels, it is likely that a jet from a $t \rightarrow b(W \rightarrow q\bar{q})$ decay is missed, while in the 2-lepton channel it is likely that an ISR or FSR jet is selected. This is the reason for not correlating, between these two sets of lepton channels, the systematic uncertainty attached to the 3-to-2 jet ratio for the $t\bar{t}$
background. Another example is the $\Delta \phi$ reweighting in the $W+\text{jets}$ processes, which is derived in the 0-tag sample and applied to the $Wc\ell$ and $Wl$ backgrounds. As explained in section 7, this reweighting is not applied to the $Wcc$ and $Wb$ backgrounds but, in the absence of further information, an uncertainty is assessed for the $\Delta \phi$ distributions of the $Wcc$ and $Wb$ backgrounds, uncorrelated with the uncertainty applied to the $Wc\ell$ and $Wl$ backgrounds. Altogether, the fit has to handle almost 170 NPs, with roughly half of those being of experimental origin.

The fit uses templates constructed from the predicted yields for the signal and the various backgrounds in the bins of the input distribution in each region. The systematic uncertainties are encoded in templates of variations relative to the nominal template for each up-and-down ($\pm 1\sigma$) variation. The limited size of the MC samples for some simulated background processes in some regions can cause large local fluctuations in templates of systematic variations. When the impact of a systematic variation translates into a reweighting of the nominal template, no statistical fluctuations are expected beyond those already present in the nominal template. This is the case, for instance, for the $b$-tagging uncertainties. For those, no specific action is taken. On the other hand, when a systematic variation may introduce changes in the events selected, as is the case for instance with the JES uncertainties, additional statistical fluctuations may be introduced, which affect the templates of systematic variations. In such cases, a smoothing procedure is applied to each systematic-variation template in each region. Bins are merged based on the constraints that the statistical uncertainty in each bin should be less than 5% and that the shapes of the systematic-variation templates remain physical: monotonous for a BDT$_{VH}$ discriminant, and with at most one local extremum for a dijet mass.

Altogether, given the number of regions and NPs, the number of systematic-variation template pairs ($\pm 1\sigma$) is close to twenty thousand, which renders the fits highly time consuming. To address this issue, systematic uncertainties that have a negligible impact on the final results are pruned away, region by region. A normalisation (shape) uncertainty is dropped if the associated template variation is below 0.5% (below 0.5% in all bins). Additional pruning criteria are applied to regions where the signal contribution is less than 2% of the total background and where the systematic variations impact the total background prediction by less than 0.5%. Furthermore, shape uncertainties are dropped if the up- and down-varied shapes are more similar to each other than to the nominal shape. This is only done for those systematic uncertainties where opposite-sign variations are expected. This procedure reduces the number of systematic-variation templates by a factor of two.

The behaviour of the global fit is evaluated by a number of checks, including how much each NP is pulled away from its nominal value, how much its uncertainty is reduced with respect to its nominal uncertainty, and which correlations develop between initially uncorrelated systematic uncertainties. To assess these effects, comparisons are made between the expectations from simulation and the observations in the data. When differences arise, their source is investigated, and this leads in a number of cases to uncorrelating further systematic uncertainties by means of additional NPs. This is to prevent a constraint from being propagated from one kinematic region to another if this is not considered well moti-
vated. The stability of the results is also tested by performing fits for each lepton channel independently, which can also help to identify from which region each constraint originates.

It is particularly useful to understand which systematic uncertainties have the largest impact on the final results, and therefore should be considered with greater care. For this purpose, a so-called ranking of the NPs is established. For each systematic uncertainty, the fit is performed again with the corresponding NP fixed to its fitted value, $\hat{\theta}$, shifted up or down by its fitted uncertainty, with all the other parameters allowed to vary so as to take properly into account the correlations between systematic uncertainties. The magnitude of the shift in the fitted signal strength $\hat{\mu}$ is a measure of the observed impact of the considered NP. The same procedure is repeated, using the nominal values of the NP and of its associated uncertainty to provide its expected impact. To reduce the computation time and therefore to enable more detailed fit studies, some of the NPs which have a negligible effect on the expected fitted uncertainty on $\hat{\mu}$ are dropped: those associated with the muon momentum scale and resolution and with the electron energy resolution; one of those associated with the jet energy scale; and those associated with the quark-gluon composition of the backgrounds, which turn out to be fully correlated with those associated with the difference in energy response between quark and gluon jets. The ranking of the systematic uncertainties obtained with the MVA applied to the 8 TeV data is shown in figure 16 with the NPs ordered by decreasing post-fit impact on $\hat{\mu}$. The five systematic uncertainties with the largest impact are, in descending order, those: on the dijet-mass shape for the $W_{bb}$ and $W_{cc}$ backgrounds for $p_T^W > 120$ GeV; on the $W_{bl}/W_{bb}$ normalisation ratio for $p_T^W > 120$ GeV; on the $W_{bb}$ background normalisation; on the $p_T^W$ shape in the 3-jet category for the $W + hf$ background; and on the signal acceptance due to the parton-shower modelling.

Since the same data sample is used for both the dijet-mass analysis and the MVA, the consistency of the two final results, i.e., the two fitted signal strengths, is assessed using the “bootstrap” method [93]. A large number of event samples are randomly extracted from the simulated samples, with the signal strength $\mu$ set to unity, the SM value. Each of them is representative of the integrated luminosity used for the data analysis in terms of expected yields as well as of associated Poisson fluctuations. Each of these event samples is subjected to both the dijet-mass analysis and the MVA, thus allowing the two fitted $\hat{\mu}$ values to be compared and their statistical correlation to be extracted. At the same time, the expected distributions of $\hat{\mu}$ and of its uncertainty are determined for both the dijet-mass analysis and the MVA.

9.3 Cross checks using diboson production

Diboson production with a $Z$ boson decaying to a pair of $b$-quarks and produced in association with either a $W$ or $Z$ boson has a signature very similar to the one considered in this analysis, but with a softer $p_T^{bb}$ spectrum and with a $m_{bb}$ distribution peaking at lower values. The cross section is about five times larger than for the SM Higgs boson with a mass of 125 GeV. Diboson production is therefore used as a validation of the analysis procedure. For the dijet-mass analysis, the binning of the transformed $m_{bb}$ distribution is reoptimised for the $Z$ boson mass. For the MVA, the BDTs are retrained to discriminate
Figure 16. Impact of systematic uncertainties on the fitted signal-strength parameter $\hat{\mu}$ for the MVA applied to the 8 TeV data. The systematic uncertainties are listed in decreasing order of their impact on $\hat{\mu}$ on the y-axis. The boxes show the variations of $\hat{\mu}$, referring to the top x-axis, when fixing the corresponding individual nuisance parameter $\theta$ to its post-fit value $\hat{\theta}$ modified upwards or downwards by its post-fit uncertainty, and repeating the fit as explained in the text. The hatched and open areas correspond to the upwards and downwards variations, respectively. The filled circles, referring to the bottom x-axis, show the deviations of the fitted nuisance parameters $\theta$ from their nominal values $\theta_0$, expressed in terms of standard deviations with respect to their nominal uncertainties $\Delta \theta$. The associated error bars show the post-fit uncertainties of the nuisance parameters, relative to their nominal uncertainties. The open circles with their error bars, also referring to the bottom x-axis, show the fitted values and uncertainties of the normalisation parameters that are freely floating in the fit. The normalisation parameters have a pre-fit value of one. As explained in section 8.1, the jet energy scale and b-tagging uncertainties are decomposed into uncorrelated components; the labels 1 and 4 refer to such components.
the diboson signal from all backgrounds (including the Higgs boson). So-called "VZ fits" are performed, where the normalisation of the diboson contributions is allowed to vary with a multiplicative scale factor $\mu_{VZ}$ with respect to the SM expectation, except for the small contribution from WW production, which is treated as a background and constrained within its uncertainty. A SM Higgs boson with $m_H = 125$ GeV is included as a background, with a production cross section at the SM value with an uncertainty of 50%. Distributions of the BDT$_{VZ}$ discriminants of the MVA are shown in figure 17 for 2-tag signal regions with $p_T^V > 120$ GeV in the 2-jet category of the 0-, 1- and 2-lepton channels.

As an additional check, fits are also performed with both the diboson and Higgs boson signal-strength parameters $\mu_{VZ}$ and $\mu$ left freely floating, to study the correlation between the two strength parameters. The fits in the dijet-mass analysis use the $m_{bb}$ distributions with binning optimised for a Higgs boson mass of 125 GeV. The fits in the MVA use BDTs trained for that same mass, as well as the associated optimised binnings.
Figure 17. Distribution of the BDT\(_{VZ}\) discriminant used in the \(VZ\) fit observed in data (points with error bars) and expected (histograms) for the 2-jet 2-tag signal regions with \(p_T > 120\) GeV in the Medium (MM) and Tight (TT) categories combined of (a) the 0-lepton channel, (b) the 1-lepton channel, (c) the 2-lepton channel. The \(VZ\) signal and background contributions are shown as filled histograms after the global \(VZ\) fit of the MVA, except for the Higgs boson contribution (\(m_H = 125\) GeV). The latter is shown as expected from the SM (indicated as \(\mu = 1.0\)) both as a filled histogram on top of the sum of the fitted \(VZ\) signal and other backgrounds, and unstacked as an unfilled histogram scaled by the factor indicated in the legend. The expected \(VZ\) contribution is shown in a similar way as an unfilled dotted histogram. The dashed histogram shows the sum of the total \(VZ\) signal and background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty on the fitted \(VZ\) signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted \(VZ\) signal and background is shown in the lower panel.
10 Analysis of the 7 TeV data

For the 7 TeV dataset, only a dijet-mass analysis is performed. It is similar but not identical to the corresponding analysis for the 8 TeV data, since some of the object reconstruction tools, such as the simultaneous use of multiple $b$-tagging operating points, are not available for the 7 TeV data. In this section, the main differences between the two analyses are summarised.

10.1 Object reconstruction

The three categories of electrons are selected according to the loose, medium, and tight criteria defined in ref. [53]. The transverse energy threshold for loose electrons is set at 10 GeV, instead of 7 GeV. For tight electrons and muons, the calorimeter isolation requirement is loosened from 4% to 7%. The procedure used to avoid double-counting of reconstructed muon and jet objects removes muons separated by $\Delta R < 0.4$ from any jet, irrespective of the multiplicity of tracks associated with the jet. For jets, the global sequential calibration is not used and the requirement on the fraction of track $p_T$ carried by tracks originating from the primary vertex is raised from 50% to 75%. The $b$-tagging algorithm used is MV1 [94–97] instead of MV1c, with a single operating point to define $b$-tagged jets corresponding to an efficiency of 70%.

10.2 Event selection

The selection criteria are those used in the dijet-mass analysis of the 8 TeV data, with the following differences. With only one $b$-tagging operating point, a single 2-tag category is defined. In the 0-lepton channel, the 100–120 GeV $p_T^{V}$ interval is not used, and the criterion for $\sum p_T^{\text{jett}}$ is not applied. In the 1-muon sub-channel, the $E_T^{\text{miss}}$ trigger is used only in the 2-jet 2-tag category for $p_T^{W} > 160$ GeV, and the events selected only by the $E_T^{\text{miss}}$ trigger constitute distinct signal regions. In the 1-lepton channel, $m_T^{W} > 40$ GeV is required for $p_T^{W} < 160$ GeV; there is no requirement on $H_T$, but $E_T^{\text{miss}} > 25$ GeV is imposed for $p_T^{W} < 200$ GeV. In the 2-lepton channel, no kinematic fit is performed. Different lepton-flavour events are used to define a 2-tag $t\bar{t}$-dominated $e-\mu$ control region in the 2-lepton channel; the region is defined to be inclusive in jet multiplicity ($\geq 2$).

10.3 Background composition and modelling

The templates used to model the MJ background in the 1-lepton channel are obtained by inversion of the track-based isolation criterion, and the normalisations are performed on the $m_T^{W}$ and $E_T^{\text{miss}}$ distributions in the electron and muon sub-channels, respectively. Corrections to the simulation of the $V+\text{jet}$ backgrounds are determined in the 1- and 2-lepton 0-tag samples inclusively in $p_T^{V}$, and applied as $\Delta \phi(\text{jet}_1, \text{jet}_2)$ reweightings to the $W+\text{jet}$ and $Z+\text{jet}$ components in all channels. Selected dijet-mass distributions showing the background composition in various analysis regions are shown in figure 18.
10.4 Systematic uncertainties

The differences with respect to the 8 TeV data analysis arise mainly from experimental systematic uncertainties. Many of them are evaluated using independent data samples (7 TeV data vs. 8 TeV data), e.g., $E_T^{\text{miss}}$ trigger efficiencies or JES. Others refer to different identification algorithms, e.g., electron identification or $b$-tagging. The uncertainty on the integrated luminosity is 1.8% for the 2011 dataset [20].

The uncertainties affecting the signal and background simulation are estimated in a similar way as for the 8 TeV data, i.e., from comparisons between the baseline and alternative generators. For $V$+jets, the $Vbc$ and $Vbb$ backgrounds are merged into a single component. For dibosons, the baseline generator is HERWIG instead of PYTHIA8; systematic uncertainties on the 3-to-2-jet ratios and on the $p_T^V$ distributions are estimated at generator level for the different diboson processes by comparison with MCFM at NLO. For the signal, the $gg \rightarrow ZH$ samples are generated with PYTHIA8 instead of POWHEG; for all processes, the alternative generators used are PYTHIA6 and HERWIG.

Due to these differences, and because the phase space within which the systematic uncertainties are evaluated is more restricted than for the MVA applied to the 8 TeV data, all systematic uncertainties, except for the theoretical uncertainties on the signal, are treated as uncorrelated between the analyses of the 7 TeV and 8 TeV data in the global fit to the combined dataset, in which the MVA is used for the 8 TeV data.

10.5 Statistical procedure

The inputs to the likelihood fits are the $m_{bb}$ distributions (not transformed) in the 28 $p_T^V$ intervals of the 2-tag signal regions. Additional inputs are the event yields in the five $p_T^V$ intervals of the 2-tag $e$–$\mu$ control region and the 26 $p_T^V$ intervals of the 1-tag control regions. For the $t\bar{t}$ background, a single floating normalisation is determined by the global fit, instead of one in each of the 0-, 1-, and 2-lepton channels. In addition to the other floating normalisations mentioned for the 8 TeV data analysis, the MJ background normalisation is also left freely floating in all regions of the 1-lepton channel, except in the 2-tag 3-jet regions where the statistical power of the data is not sufficient to provide a reliable constraint. In these regions, an uncertainty of 30% is assigned to the MJ background normalisation, using a method similar to what is done for the analysis of the 8 TeV data.
Figure 18. The dijet-mass distributions observed in the 7 TeV data (points with error bars) and expected (histograms) for the 2-jet 2-tag signal regions with the $p_T > 120$ GeV intervals combined; (a) 0-lepton channel, (b) 1-lepton channel, and (c) 2-lepton channel. The background contributions after the global fit of the dijet-mass analysis are shown as filled histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled histogram on top of the fitted backgrounds, as expected from the SM (indicated as $\mu = 1.0$), and, unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The size of the combined statistical and systematic uncertainty on the sum of the signal and fitted background is indicated by the hatched band. The ratio of the data to the sum of the signal and fitted background is shown in the lower panel.
11 Results

As explained in section 9, the results are obtained from maximum-likelihood fits to the data, where the inputs are the distributions of final discriminants in the 2-tag signal regions and the $MV_{1c}$ distributions of the $b$-tagged jet in the 1-tag control regions, with nuisance parameters either floating or constrained by priors. The final discriminants are the transformed $m_{bb}$ for the dijet-mass analysis and the BDT$_{VH}$ discriminants for the MVA. Results are extracted independently for the dijet-mass and multivariate analyses. Since the MVA has better expected sensitivity to a Higgs boson signal, it is used for the nominal results, while the dijet-mass analysis provides a cross-check (cf. section 11.2). For the 7 TeV data, however, only a dijet-mass analysis is performed. Unless otherwise specified, all results refer to a Higgs boson mass of 125 GeV.

In the following, the fitted signal-strength parameters are simply denoted $\mu$ and $\mu_{VZ}$, rather than $\hat{\mu}$ and $\hat{\mu}_{VZ}$.

11.1 Nominal results

The nominal results are obtained from global fits using the MVA for the 8 TeV data and the dijet-mass analysis of the 7 TeV data.

Distributions of the BDT$_{VH}$ discriminant and of $MV_{1c}$, with background normalisations and nuisance parameters adjusted by the global fit to the 8 TeV data were already presented in section 7.4. Dijet-mass distributions in the 7 TeV data analysis were shown in section 10. Agreement between data and estimated background is observed within the uncertainties shown by the hatched bands.

Figure 19 shows the 95% CL upper limits on the cross section times branching ratio for $pp \rightarrow (W/Z)(H \rightarrow b\bar{b})$ in the Higgs boson mass range 110–140 GeV. The observed limit for $m_H = 125$ GeV is 1.2 times the SM value, to be compared to an expected limit, in the absence of signal, of 0.8. For the 8 TeV (7 TeV) data only, the observed and expected limits are 1.4 (2.3) and 0.8 (3.2), respectively.

The probability $p_0$ of obtaining from background alone a result at least as signal-like as the observation is 8% for a tested Higgs boson mass of 125 GeV; in the presence of a Higgs boson with that mass and the SM signal strength, the expected $p_0$ value is 0.5%. This corresponds to an excess observed with a significance of 1.4$\sigma$, to be compared to an expectation of 2.6$\sigma$. For the 8 TeV data alone, the observed and expected levels of significance are 1.7$\sigma$ and 2.5$\sigma$, respectively. For the 7 TeV data alone, the expected significance is 0.7$\sigma$ and there is a deficit rather than an excess in the data, as can be seen in figure 18. Figure 20 shows the $p_0$ values in the mass range 110–140 GeV, as obtained for the 7 TeV and 8 TeV combined dataset.

The fitted $\mu$ values for $m_H = 125$ GeV are shown in figure 21 for the 7 TeV, 8 TeV and combined datasets. With all lepton channels and data-taking periods combined, the fitted value of the signal-strength parameter is\textsuperscript{11} $\mu = 0.51 \pm 0.31\text{(stat.)} \pm 0.24\text{(syst.)}$. For the 8 TeV

\textsuperscript{11} The uncertainties of the normalisations of the floating backgrounds are included in the systematic uncertainties; their contribution is 0.07.
Figure 19. Observed (solid) and expected 95% CL cross-section upper limits, normalised to the SM Higgs boson production cross section, as a function of $m_H$ for all channels and data-taking periods combined, as obtained using the dijet-mass analysis for the 7 TeV dataset and BDTs trained at each individual mass for the 8 TeV dataset. The expected upper limit is given for the background-only hypothesis (dashed) and with the injection of a SM Higgs boson signal at a mass of 125 GeV (dotted). The dark and light shaded bands represent the 1σ and 2σ ranges of the expectation in the absence of a signal. For all curves shown, the results obtained at the tested masses are linearly interpolated.

Figure 20. Observed (solid) and expected $p_0$ values as a function of $m_H$ for all channels and data-taking periods combined, as obtained using the dijet-mass analysis for the 7 TeV dataset and BDTs trained at each individual mass for the 8 TeV dataset. The expected $p_0$ values are given for the background-only hypothesis in the presence of a SM Higgs boson: for the dashed curve the Higgs boson mass corresponds to each tested mass point in turn; for the dotted curve the Higgs boson mass is 125 GeV. For all curves shown, the results obtained at the tested masses are linearly interpolated.
data, the fitted value of the signal-strength parameter is $\mu = 0.52 \pm 0.32 \text{(stat.)} \pm 0.24 \text{(syst.)}$. For the 7 TeV data, it is $\mu = -1.6 \pm 1.2 \text{(stat.)} \pm 0.9 \text{(syst.)}$. 

For a Higgs boson with a mass of 125.36 GeV, as measured by ATLAS [98], the signal-strength parameter is $\mu = 0.52 \pm 0.32 \text{(stat.)} \pm 0.24 \text{(syst.)}$. 

Fits are also performed where the signal strengths are floated independently for (i) the $WH$ and $ZH$ production processes, or (ii) the three lepton channels. The results of these fits are shown in figures 22 and 23 respectively. The consistency of the fitted signal strengths in the $WH$ and $ZH$ processes is at the level of 20%. For the lepton channels, the consistency between the three fitted signal strengths is at the level of 72% for the 7 TeV data, and of 8% for the 8 TeV data. The low values of the fitted signal strengths for the $ZH$ process and in the 0-lepton channel are associated with the data deficit observed in the most sensitive bins of the BDT$_{VH}$ discriminant in the 0-lepton channel, shown in figure 12(a).

Figure 24 shows the data, background and signal yields, where the final-discriminant bins in all signal regions are combined into bins of $\log(S/B)$, separately for the 7 and 8 TeV datasets. Here, $S$ is the expected signal yield and $B$ is the fitted background yield. Details of the fitted values of the signal and of the various background components are provided in table 8.

11.2 Cross-check with the dijet-mass analysis

The distributions of $m_{bb}$ in the dijet-mass analysis, with background normalisations and nuisance parameters adjusted by the global fit to the 8 TeV data were already presented in section 7.3. Agreement between data and estimated background is observed within the uncertainties shown by the hatched bands.
Figure 22. The fitted values of the Higgs boson signal-strength parameter $\mu$ for $m_H = 125$ GeV for the $WH$ and $ZH$ processes and the combination of the $WH$ and $ZH$ processes, with the 7 and 8 TeV datasets combined. The individual $\mu$ values for the ($W/Z$)H processes are obtained from a simultaneous fit with the signal strength for each of the $WH$ and $ZH$ processes floating independently.

Figure 23. The fitted values of the Higgs boson signal-strength parameter $\mu$ for $m_H = 125$ GeV for the 0-, 1- and 2-lepton channels and the combination of the three channels, with the 7 and 8 TeV datasets combined. The individual $\mu$ values for the lepton channels are obtained from a simultaneous fit with the signal strength for each of the lepton channels floating independently.
Figure 24. Event yields as a function of log(S/B) for data, background and Higgs boson signal with $m_H = 125$ GeV for the (a) 8 TeV data and (b) 7 TeV data. Final-discriminant bins in all signal regions are combined into bins of log(S/B). The signal $S$ and background $B$ yields are expected and fitted, respectively. The Higgs boson signal contribution is shown as expected for the SM cross section (indicated as $\mu = 1.0$). The pull of the data with respect to the background-only prediction is also shown with statistical uncertainties only. The full line indicates the pull of the prediction for signal ($\mu = 1.0$) and background with respect to the background-only prediction.

Table 8. The numbers of expected signal and fitted background events and the observed numbers of events after MVA selection in the bins of figure 24(a). These numbers are for both the 1-tag and 2-tag events in the 8 TeV dataset, corresponding to an integrated luminosity of 20.3 fb$^{-1}$. 

<table>
<thead>
<tr>
<th>Process</th>
<th>Bin 1</th>
<th>Bin 2</th>
<th>Bin 3</th>
<th>Bin 4</th>
<th>Bin 5</th>
<th>Bin 6</th>
<th>Bin 7</th>
<th>Bin 8</th>
<th>Bin 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>368550</td>
<td>141166</td>
<td>111865</td>
<td>20740</td>
<td>5538</td>
<td>2245</td>
<td>382</td>
<td>41</td>
<td>4</td>
</tr>
<tr>
<td>Signal</td>
<td>29</td>
<td>43</td>
<td>96</td>
<td>57</td>
<td>58</td>
<td>62</td>
<td>32</td>
<td>10.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Background</td>
<td>368802</td>
<td>140846</td>
<td>111831</td>
<td>20722</td>
<td>5467</td>
<td>2189</td>
<td>364</td>
<td>37.9</td>
<td>3.4</td>
</tr>
<tr>
<td>$S/B$</td>
<td>$8 \times 10^{-5}$</td>
<td>0.0003</td>
<td>0.0009</td>
<td>0.003</td>
<td>0.01</td>
<td>0.09</td>
<td>0.3</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>$W$+hf</td>
<td>14584</td>
<td>10626</td>
<td>15297</td>
<td>1948</td>
<td>618</td>
<td>250</td>
<td>45</td>
<td>8.2</td>
<td>0.7</td>
</tr>
<tr>
<td>$Wcl$</td>
<td>96282</td>
<td>30184</td>
<td>15227</td>
<td>1286</td>
<td>239</td>
<td>47</td>
<td>4.2</td>
<td>0.2</td>
<td>0.005</td>
</tr>
<tr>
<td>$Wl$</td>
<td>125676</td>
<td>14961</td>
<td>3722</td>
<td>588</td>
<td>107</td>
<td>16</td>
<td>1.3</td>
<td>0.03</td>
<td>0.001</td>
</tr>
<tr>
<td>$Z$+hf</td>
<td>10758</td>
<td>14167</td>
<td>21684</td>
<td>7458</td>
<td>1178</td>
<td>577</td>
<td>130</td>
<td>14.8</td>
<td>2.2</td>
</tr>
<tr>
<td>$Zcl$</td>
<td>13876</td>
<td>11048</td>
<td>4419</td>
<td>941</td>
<td>61</td>
<td>22</td>
<td>2.1</td>
<td>0.1</td>
<td>0.008</td>
</tr>
<tr>
<td>$Zl$</td>
<td>49750</td>
<td>18061</td>
<td>3044</td>
<td>537</td>
<td>48</td>
<td>15</td>
<td>1</td>
<td>0.05</td>
<td>0.004</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>30539</td>
<td>24824</td>
<td>26729</td>
<td>5595</td>
<td>2238</td>
<td>922</td>
<td>137</td>
<td>10</td>
<td>0.3</td>
</tr>
<tr>
<td>Single top</td>
<td>10356</td>
<td>9492</td>
<td>14279</td>
<td>1494</td>
<td>688</td>
<td>252</td>
<td>31</td>
<td>2.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Diboson</td>
<td>4378</td>
<td>1831</td>
<td>1247</td>
<td>474</td>
<td>186</td>
<td>62</td>
<td>9.7</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Multijet</td>
<td>12603</td>
<td>5650</td>
<td>6184</td>
<td>400</td>
<td>103</td>
<td>26</td>
<td>3</td>
<td>0.9</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 25. The distribution of \( m_{bb} \) in data after subtraction of all backgrounds except for the diboson processes, as obtained with the dijet-mass analysis for the (a) 8 TeV and (b) 7 TeV data. The contributions from all lepton channels, \( p_T \) intervals, number-of-jets and 2-tag \( b \)-tagging categories are summed weighted by their respective values of the ratio of expected Higgs boson signal to fitted background. The contribution of the associated \( WH \) and \( ZH \) production of a SM Higgs boson with \( m_H = 125 \text{ GeV} \) is shown as expected for the SM cross section (indicated as \( \mu = 1.0 \)). The size of the combined statistical and systematic uncertainty on the fitted background is indicated by the hatched band.

In the dijet-mass analysis, a \( \mu \) value of \( 1.23 \pm 0.44 \text{ (stat.)} \pm 0.41 \text{ (syst.)} \) is obtained for the 8 TeV dataset. The consistency of the results of the three lepton channels is at the level of 8%. Using the “bootstrap” method mentioned in section 9.2, the results for the 8 TeV data with the dijet-mass analysis and with the MVA are expected to be 67% correlated, and the observed results are found to be statistically consistent at the level of 8%. The observed significance in the dijet-mass analysis is 2.2\( \sigma \). The expected significance is 1.9\( \sigma \), to be compared to 2.5\( \sigma \) for the MVA, which is the reason for choosing the MVA for the nominal results.

Figure 25 shows the \( m_{bb} \) distribution in data after subtraction of all backgrounds except for diboson production for the 7 and 8 TeV data, as obtained with the dijet-mass analysis. In this figure, the contributions of all 2-tag signal regions in all channels are summed weighted by their respective ratios of expected Higgs boson signal to fitted background. The \( VZ \) contribution is clearly seen, located at the expected \( Z \) mass. The Higgs boson signal contribution is shown as expected for the SM cross section.

11.3 Cross-check with the diboson analysis

To validate the analysis procedures, \( VZ \) fits are performed, the technical details of which were discussed in section 9.3.
The measured signal strength for the 8 TeV dataset with the MVA is $\mu_{VZ} = 0.77 \pm 0.10^{\text{stat.}} \pm 0.15^{\text{syst.}}$. This result is consistent with the observations already made on figure 25. The signal strengths obtained for the three lepton channels are consistent at the 85% level. In the dijet-mass analysis at 8 TeV, a $\mu_{VZ}$ value of $0.79 \pm 0.11^{\text{stat.}} \pm 0.16^{\text{syst.}}$ is obtained. The correlation of the systematic uncertainties on $\mu_{VZ}$ and $\mu$ is 35% in the MVA and 67% in the dijet-mass analysis.

Fits are performed with the same final discriminants as used to obtain the results for the Higgs boson based on the 8 TeV dataset, but with both the $VZ$ and Higgs boson signal-strength parameters $\mu_{VZ}$ and $\mu$ left freely floating. The result for the Higgs boson signal strength is unchanged from the nominal result, and the statistical correlation between the two signal-strength parameters is found to be $-3\%$ in the MVA and 9% in the dijet-mass analysis. The main reason for these low correlations is the different shape of the $p_T^V$ distributions for $VZ$ and for the Higgs boson signal, the $p_T^V$ variable being used by both the MVA and the dijet-mass analysis. The yield tables in the appendix show that the ratio of the diboson contribution to that of the Higgs boson is indeed smaller in the higher $p_T^V$ interval than in the lower one. The additional variables input to the BDT provide further separation in the MVA, leading to a very small diboson contribution in the most significant bins of the BDT$_V$H discriminant, as can be seen in table 8.

A value of $\mu_{VZ} = 0.50 \pm 0.30^{\text{(stat.)}} \pm 0.38^{\text{(syst.)}}$ is obtained for the 7 TeV dataset. The signal strength obtained for the combined 7 and 8 TeV dataset is $0.74 \pm 0.09^{\text{(stat.)}} \pm 0.14^{\text{(syst.)}}$. The $VZ$ signal is observed with a significance of $4.9\sigma$, to be compared to an expected significance of $6.3\sigma$. The fitted $\mu_{VZ}$ values are shown in figure 26 for the 7 TeV, 8 TeV and combined datasets, and for the three lepton channels separately for the combined dataset, all with the MVA used for the 8 TeV data. A measurement of $VZ$ production in $pp$ collisions at $\sqrt{s} = 8$ TeV in final states with $b$-tagged jets was recently reported by the CMS Collaboration [99].
Figure 26. The fitted values of the diboson signal strength $\mu_{VZ}$ for (a) the 7 TeV, 8 TeV and combined datasets, and (b) for the three lepton channels separately and combined, for the combined dataset. The MVA is used for the 8 TeV data. The individual $\mu_{VZ}$ values for the lepton channels are obtained from a simultaneous fit with the signal strength for each floating independently.

12 Summary

A search for the Standard Model Higgs boson produced in association with a $W$ or $Z$ boson and decaying into $b\bar{b}$ has been presented. The ($W/Z$) decay channels considered are $W \to \ell\nu$, $Z \to \ell\ell$ and $Z \to \nu\nu$. The dataset corresponds to integrated luminosities of 4.7 fb$^{-1}$ and 20.3 fb$^{-1}$ from $pp$ collisions at 7 TeV and 8 TeV, respectively, recorded by the ATLAS experiment during Run 1 of the LHC.

The analysis is carried out in event categories based on the numbers of leptons, jets, and jets tagged as originating from $b$-quark fragmentation, and on the transverse momentum of the vector-boson candidate. A multivariate analysis provides the nominal results. An alternative analysis using invariant-mass distributions of the Higgs boson candidates leads to consistent results.

For a Higgs boson mass of 125.36 GeV, the observed (expected) deviation from the background-only hypothesis corresponds to a significance of 1.4 (2.6) standard deviations and the ratio of the measured signal yield to the Standard Model expectation is found to be $\mu = 0.52 \pm 0.32$(stat.) $\pm 0.24$(syst.). The analysis procedure is validated by a measurement of the yield of ($W/Z$)$Z$ production with $Z \to b\bar{b}$, from which the ratio of the observed signal yield to the Standard Model expectation is found to be $0.74 \pm 0.09$(stat.) $\pm 0.14$(syst.).

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST...
The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (U.K.) and BNL (U.S.A.) and in the Tier-2 facilities worldwide.

A Tables of event yields

The event yields in each category for the multivariate analysis are shown in tables 9–11.
<table>
<thead>
<tr>
<th>Sample $p_T^{V}$</th>
<th>1-tag $100 - 120 \text{ GeV}$</th>
<th>&gt; $120 \text{ GeV}$</th>
<th>LL-tag $100 - 120 \text{ GeV}$</th>
<th>&gt; $120 \text{ GeV}$</th>
<th>MM-tag $100 - 120 \text{ GeV}$</th>
<th>&gt; $120 \text{ GeV}$</th>
<th>MM+TT-tag $100 - 120 \text{ GeV}$</th>
<th>&gt; $120 \text{ GeV}$</th>
<th>TT-tag $100 - 120 \text{ GeV}$</th>
<th>&gt; $120 \text{ GeV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VH$</td>
<td>7.9</td>
<td>23</td>
<td>2.2</td>
<td>6.9</td>
<td>3.5</td>
<td>23</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$WZ, ZZ, WW$</td>
<td>235</td>
<td>635</td>
<td>18</td>
<td>49</td>
<td>14.7</td>
<td>81</td>
<td>13.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>840</td>
<td>1520</td>
<td>114</td>
<td>183</td>
<td>129</td>
<td>332</td>
<td>116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single top</td>
<td>531</td>
<td>704</td>
<td>40</td>
<td>56</td>
<td>32.6</td>
<td>66</td>
<td>22.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Wl$</td>
<td>5470</td>
<td>7100</td>
<td>159</td>
<td>206</td>
<td>14.4</td>
<td>16</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Wcl$</td>
<td>2230</td>
<td>3710</td>
<td>106</td>
<td>159</td>
<td>23.4</td>
<td>27</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W+hf$</td>
<td>762</td>
<td>1520</td>
<td>54</td>
<td>124</td>
<td>33.6</td>
<td>128</td>
<td>21.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Zl$</td>
<td>3890</td>
<td>10750</td>
<td>96</td>
<td>272</td>
<td>6.1</td>
<td>17</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Zcl$</td>
<td>1590</td>
<td>3990</td>
<td>59</td>
<td>162</td>
<td>9.4</td>
<td>24</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z+hf$</td>
<td>2550</td>
<td>6510</td>
<td>225</td>
<td>607</td>
<td>186</td>
<td>876</td>
<td>151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18340 ± 150</strong></td>
<td><strong>36890 ± 200</strong></td>
<td><strong>886 ± 17</strong></td>
<td><strong>1841 ± 25</strong></td>
<td><strong>458.5 ± 9.8</strong></td>
<td><strong>1599 ± 28</strong></td>
<td><strong>333.2 ± 8.9</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>18343</td>
<td>36903</td>
<td>887</td>
<td>1860</td>
<td>477</td>
<td>1592</td>
<td>306</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VH$</td>
<td>–</td>
<td>8</td>
<td>–</td>
<td>2</td>
<td>–</td>
<td>7</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$WZ, ZZ, WW$</td>
<td>–</td>
<td>260</td>
<td>–</td>
<td>17</td>
<td>–</td>
<td>20</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>–</td>
<td>1670</td>
<td>–</td>
<td>186</td>
<td>–</td>
<td>315</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Wl$</td>
<td>–</td>
<td>2280</td>
<td>–</td>
<td>59</td>
<td>–</td>
<td>4.3</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Wcl$</td>
<td>–</td>
<td>1240</td>
<td>–</td>
<td>53</td>
<td>–</td>
<td>8.9</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W+hf$</td>
<td>–</td>
<td>750</td>
<td>–</td>
<td>60</td>
<td>–</td>
<td>62</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Zl$</td>
<td>–</td>
<td>3190</td>
<td>–</td>
<td>79</td>
<td>–</td>
<td>4.5</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Zcl$</td>
<td>–</td>
<td>1620</td>
<td>–</td>
<td>65</td>
<td>–</td>
<td>9.8</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Z+hf$</td>
<td>–</td>
<td>1890</td>
<td>–</td>
<td>170</td>
<td>–</td>
<td>259</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>–</td>
<td><strong>13310 ± 100</strong></td>
<td>–</td>
<td><strong>718 ± 12</strong></td>
<td>–</td>
<td><strong>719 ± 17</strong></td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>–</td>
<td>13344</td>
<td>–</td>
<td>657</td>
<td>–</td>
<td>710</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. The expected signal and fitted background yields for each category of the 0-lepton channel after the full selection of the multivariate analysis. The $m_{bb}$ distribution is used in the $100 < p_T^{V} < 120 \text{ GeV}$ interval. The background yields are normalised by the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties.
Table 10. The expected signal and fitted background yields for each category of the 1-lepton channel after the full selection of the multivariate analysis. The background yields are normalised by the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties.
Table 11. The expected signal and fitted background yields for each category of the 2-lepton channel after the full selection of the multivariate analysis. The background yields are normalised by the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties.

<table>
<thead>
<tr>
<th>Sample</th>
<th>1-tag</th>
<th>LL-tag</th>
<th>MM+TT-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_T^V$</td>
<td>$&lt; 120 \text{ GeV}$</td>
<td>$&gt; 120 \text{ GeV}$</td>
</tr>
<tr>
<td></td>
<td>2-jet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VH</td>
<td>20</td>
<td>7.4</td>
<td>5.5</td>
</tr>
<tr>
<td>WZ,ZZ,WW</td>
<td>800</td>
<td>135</td>
<td>73</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>3490</td>
<td>213</td>
<td>1023</td>
</tr>
<tr>
<td>Single top</td>
<td>385</td>
<td>44</td>
<td>40</td>
</tr>
<tr>
<td>$W^+l$</td>
<td>245</td>
<td>12</td>
<td>3.7</td>
</tr>
<tr>
<td>$W^+cl$</td>
<td>166</td>
<td>18</td>
<td>5.8</td>
</tr>
<tr>
<td>$W^+hf$</td>
<td>90</td>
<td>13</td>
<td>7.5</td>
</tr>
<tr>
<td>$Z^+l$</td>
<td>29410</td>
<td>4180</td>
<td>648</td>
</tr>
<tr>
<td>$Z^+cl$</td>
<td>12130</td>
<td>1622</td>
<td>421</td>
</tr>
<tr>
<td>$Z^+hf$</td>
<td>21090</td>
<td>2853</td>
<td>1916</td>
</tr>
<tr>
<td>Total</td>
<td>69360</td>
<td>± 280</td>
<td>9123 ± 86</td>
</tr>
<tr>
<td>Data</td>
<td>69313</td>
<td>9150</td>
<td>4313</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>3-jet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>VH</td>
<td>7.4</td>
</tr>
<tr>
<td>WZ,ZZ,WW</td>
<td>347</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>2170</td>
</tr>
<tr>
<td>Single top</td>
<td>144</td>
</tr>
<tr>
<td>$W^+l$</td>
<td>76</td>
</tr>
<tr>
<td>$W^+cl$</td>
<td>57</td>
</tr>
<tr>
<td>$W^+hf$</td>
<td>40</td>
</tr>
<tr>
<td>$Z^+l$</td>
<td>8870</td>
</tr>
<tr>
<td>$Z^+cl$</td>
<td>4650</td>
</tr>
<tr>
<td>$Z^+hf$</td>
<td>5790</td>
</tr>
<tr>
<td>Total</td>
<td>22720</td>
</tr>
<tr>
<td>Data</td>
<td>22662</td>
</tr>
</tbody>
</table>
Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 JINST 3 S08003 [SPIRE].


G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/Hz + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* 10 (2013) 083 [arXiv:1306.2542] [SPIRE].


[71] A. Höcker et al., TMVA: Toolkit for Multivariate Data Analysis, PoS(ACAT)040 [physics/0703039] [INSPHERE].


[98] ATLAS collaboration, *Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb$^{-1}$ of pp collision data*, Phys. Rev. D 90 (2014) 052004 [arXiv:1406.3827] [insPIRE].

Also at School of Physics and Engineering, Sun Yat-sen University, Guangzhou, China
Also at Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia
Also at National Research Nuclear University MEPhI, Moscow, Russia
Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
Also at Department of Physics, Oxford University, Oxford, United Kingdom
Also at Department of Physics, Nanjing University, Jiangsu, China
Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany
Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
Also at Discipline of Physics, University of KwaZulu-Natal, Durban, South Africa
Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia ∗ Deceased