Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector


DOI
10.1007/JHEP12(2017)024

Publication date
2017

Document Version
Final published version

Published in
The Journal of High Energy Physics

License
CC BY

Citation for published version (APA):
https://doi.org/10.1007/JHEP12(2017)024

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).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Evidence for the $H \rightarrow b \bar{b}$ decay with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A search for the decay of the Standard Model Higgs boson into a $b \bar{b}$ pair when produced in association with a $W$ or $Z$ boson is performed with the ATLAS detector. The analysed data, corresponding to an integrated luminosity of 36.1 fb$^{-1}$, were collected in proton-proton collisions in Run 2 of the Large Hadron Collider at a centre-of-mass energy of 13 TeV. Final states containing zero, one and two charged leptons (electrons or muons) are considered, targeting the decays $Z \rightarrow \nu \bar{\nu}$, $W \rightarrow \ell \nu$ and $Z \rightarrow \ell \ell$. For a Higgs boson mass of 125 GeV, an excess of events over the expected background from other Standard Model processes is found with an observed significance of 3.5 standard deviations, compared to an expectation of 3.0 standard deviations. This excess provides evidence for the Higgs boson decay into $b$-quarks and for its production in association with a vector boson. The combination of this result with that of the Run 1 analysis yields a ratio of the measured signal events to the Standard Model expectation equal to $0.90 \pm 0.18$(stat.)$^{+0.21}_{-0.19}$(syst.). Assuming the Standard Model production cross-section, the results are consistent with the value of the Yukawa coupling to $b$-quarks in the Standard Model.

KEYWORDS: Hadron-Hadron scattering (experiments), Higgs physics

ArXiv ePrint: 1708.03299
## Contents

1 Introduction 1

2 ATLAS detector 2

3 Dataset and simulated event samples 3

4 Object and event selection 5
   4.1 Object reconstruction 5
   4.2 Event selection and categorisation 9
      4.2.1 Zero-lepton selection 10
      4.2.2 One-lepton selection 12
      4.2.3 Two-lepton selection 13
   4.3 Selection for the dijet-mass analysis 14

5 Multivariate analysis 14

6 Estimation of the multi-jet background 15
   6.1 0-lepton channel 16
   6.2 1-lepton channel 17
   6.3 2-lepton channel 18

7 Systematic uncertainties 19
   7.1 Experimental uncertainties 19
   7.2 Simulated background uncertainties 20
   7.3 Multi-jet background uncertainties 24
   7.4 Signal uncertainties 25

8 Statistical analysis 26
   8.1 Analysis of the 13 TeV data 26
   8.2 Dijet-mass analysis 30
   8.3 Diboson analysis 30
   8.4 Combination with Run 1 data 30

9 Results 31
   9.1 Results of the SM Higgs boson search at $\sqrt{s} = 13$ TeV 31
   9.2 Results of the dijet-mass analysis 36
   9.3 Results of the diboson analysis 40
   9.4 Results of the combination with Run 1 40

10 Conclusion 43

The ATLAS collaboration 53
1 Introduction

The Higgs boson, predicted more than 50 years ago [1–4], was discovered in 2012 by the ATLAS and CMS Collaborations [5, 6], analysing the results of proton-proton (pp) collisions produced by the Large Hadron Collider (LHC) [7]. The properties of the discovered particle have been measured using the Run 1 dataset, collected at centre-of-mass energies of 7 TeV and 8 TeV, and were found to be compatible with those predicted by the Standard Model (SM) within uncertainties, typically of the order of ±20% [8–11]. The Run 2 dataset at an energy of 13 TeV provides an opportunity to increase the precision of such measurements, and to challenge theory predictions further. While analyses of Higgs bosons decaying into vector bosons are entering an era of detailed differential measurements, direct evidence for the coupling of the Higgs boson to fermions was established only via the observation of the decay into τ-leptons through the combination of ATLAS and CMS Run 1 results [11], and, more recently, through the combination of CMS Run 1 and Run 2 results [12]. Although the gluon-gluon fusion production mode provides indirect evidence for the coupling of the Higgs boson to top quarks, there is currently no direct observation of the coupling of the Higgs boson to quarks.

The decay of the SM Higgs boson into pairs of b-quarks is expected to have a branching ratio of 58% for \( m_H = 125 \) GeV [13], the largest among all decay modes. Accessing \( H \to bb \) decays is therefore crucial for constraining, under fairly general assumptions [14, 15], the overall Higgs boson decay width. At the LHC, the very large backgrounds arising from multi-jet production make an inclusive search extremely challenging. The most sensitive production modes for probing \( H \to bb \) decays are those where the Higgs boson is produced in association with a W or Z boson [16]; their leptonic decay modes lead to clean signatures that can be efficiently triggered on, while rejecting most of the multi-jet backgrounds.

Searches for a Higgs boson in the \( bb \) decay mode were conducted at the Tevatron by the CDF and D0 Collaborations. They reported an excess of events in \( VH \) associated production (where \( V \) is used to denote W or Z) in the mass range of 120 GeV to 135 GeV, with a global significance of 3.1 standard deviations, and a local significance of 2.8 standard deviations at a mass of 125 GeV [17]. ATLAS and CMS reported results from Run 1 each using approximately 25 fb\(^{-1}\) of integrated luminosity [18, 19]. Excesses of events consistent with a Higgs boson with a mass of 125 GeV were observed in \( VH \) associated production with significances of 1.4 and 2.1 standard deviations by ATLAS and CMS, respectively. Searches for the Higgs boson decay into \( bb \) have been also performed for the vector-boson fusion (VBF) [20, 21] and \( t\bar{t}H \) [22, 23] production modes, but with sensitivities smaller than for \( VH \) production. The combination of the Run 1 ATLAS and CMS analyses resulted in observed and expected significances of 2.6 and 3.7 standard deviations for the \( H \to bb \) decay mode, respectively [11].

This article reports on the search for the SM Higgs boson in the \( VH \) production mode and decaying into a \( bb \) pair with the ATLAS detector in Run 2 of the LHC, using an integrated luminosity of 36.1 fb\(^{-1}\). The Higgs boson mass is assumed to be \( m_H = 125 \) GeV, compatible with the value of \( m_H = 125.09 \pm 0.21 \) (stat.) \( \pm 0.11 \) (syst.) GeV measured by the combination of the \( H \to ZZ^* \to 4\ell \) and \( H \to \gamma\gamma \) analyses by ATLAS and CMS at the
end of Run 1 [24]. Three main signatures are explored, $ZH \rightarrow \nu \nu b\bar{b}$, $WH \rightarrow \ell \nu b\bar{b}$ and $ZH \rightarrow \ell \ell b\bar{b}$. The respective analysis categories that target these decay modes are referred to as the 0-, 1- and 2-lepton channels, based on the number of selected charged leptons. In this article, the term “lepton”, unless modified by a qualifier, refers to electron and muon. A $b$-tagging algorithm is used to identify the jets consistent with originating from a $H \rightarrow b\bar{b}$ decay. In order to maximise the sensitivity to the Higgs boson signal, a set of observables encoding information about event kinematics and topology is combined into a multivariate discriminant. A binned maximum-likelihood fit, referred to as the global likelihood fit, is applied to data simultaneously across the three channels in multiple analysis regions. The likelihood fit uses the multivariate discriminant as the main fit observable, in order to extract the signal yield and normalisations of the main backgrounds. The signal extraction method is validated with two other analyses: the dijet-mass analysis, where the signal yield is extracted using the mass of the dijet system of $b$-tagged jets as the main fit observable, and the diboson analysis, where the nominal multivariate analysis is modified to extract the $(W/Z)Z$ diboson process, with the $Z$ boson decaying into $b\bar{b}$. The combination of the results of the Higgs boson search with those of the previously published analysis of the Run 1 dataset [18] is also presented.

2 ATLAS detector

ATLAS [25] is a general-purpose particle detector covering nearly the entire solid angle\(^1\) around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon spectrometer incorporating three large superconducting toroidal magnets.

The inner tracking detector (ID or inner detector in the rest of the article), located within a 2 T axial magnetic field generated by the superconducting solenoid, is used to measure the trajectories and momenta of charged particles. The inner layers, consisting of high-granularity silicon pixel detectors, instrument a pseudorapidity range $|\eta| < 2.5$. A new innermost silicon pixel layer, the insertable B-layer [26] (IBL), was added to the detector between Run 1 and Run 2. The IBL improves the ability to identify displaced vertices and thereby significantly improves the $b$-tagging performance [27]. Silicon strip detectors covering $|\eta| < 2.5$ are located beyond the pixel detectors. Outside the strip detectors and covering $|\eta| < 2.0$, there are straw-tube tracking detectors, which also provide measurements of transition radiation that are used in electron identification.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel ($|\eta| < 1.475$) and endcap ($1.375 < |\eta| < 3.2$) high-granularity lead/liquid-argon (LAr) electromagnetic calorimeters.

---

\(^1\)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.
calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by a steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters extend the coverage to $|\eta| = 3.2$. The solid angle coverage for $|\eta|$ between 3.2 and 4.9 is completed with copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic measurements, respectively.

The outermost part of the detector is the muon spectrometer, which measures the curved trajectories of muons in the field of three large air-core toroidal magnets. High-precision tracking is performed within the range $|\eta| < 2.7$ and there are chambers for fast triggering within the range $|\eta| < 2.4$.

A two-level trigger system [28] is used to reduce the recorded data rate. The first level is a hardware implementation that makes use of only a subset of the total available information to make fast decisions to accept or reject an event, aiming to reduce the rate to approximately 100 kHz, and the second level is the software-based high-level trigger that provides the remaining rate reduction to approximately 1 kHz.

3 Dataset and simulated event samples

The data used in this analysis were collected at a centre-of-mass energy of 13 TeV during the 2015 and 2016 running periods, and correspond to integrated luminosities of $3.2 \pm 0.1 \text{ fb}^{-1}$ and $32.9 \pm 1.1 \text{ fb}^{-1}$, respectively [29]. They were collected using missing transverse momentum ($E_T^{\text{miss}}$) triggers for the 0- and 1-lepton channels and single-lepton triggers for the 1- and 2-lepton channels. Events are selected for analysis only if they are of good quality and if all the relevant detector components are known to be in good operating condition. In the combined dataset, the recorded events have an average of 25 inelastic $pp$ collisions (the collisions other than the hard scatter are referred to as pile-up).

Monte Carlo (MC) simulated events are used to model the SM background and $VH$, $H \rightarrow b\bar{b}$ signal processes. All simulated processes are normalised using the most accurate theoretical predictions currently available for their cross-sections. Data-driven methods are used to estimate the multi-jet background from strong interactions (QCD) for the 1-lepton channel, as discussed in section 6. This background is negligible in the other channels, as a result either of the high $E_T^{\text{miss}}$ requirement and dedicated selection criteria (0-lepton channel) or of the two lepton selection (2-lepton channel).

All samples of simulated events were passed through the ATLAS detector simulation [30] based on GEANT 4 [31] and are reconstructed with the standard ATLAS reconstruction software. The effects of pile-up from multiple interactions in the same and nearby bunch crossings were modelled by overlaying minimum-bias events, simulated using the soft QCD processes of PYTHIA 8.186 [32] with the A2 [33] set of tuned parameters (tune) and MSTW2008LO [34] parton distribution functions (PDF). For all samples of simulated events, except for those generated using SHERPA [35], the EvtGen v1.2.0 program [36] was used to describe the decays of bottom and charm hadrons. A summary of
Table 1. The generators used for the simulation of the signal and background processes. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant ($\alpha_S$). The acronyms ME, PS and UE stand for matrix element, parton shower and underlying event, respectively. (*) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to PDF4LHC15NLO set [37] using the internal algorithm in Powheg-Box v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$ process is normalised using the NNLO(QCD)+NLO(EW) cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ contribution.

all the generators used for the simulation of the signal and background processes is shown in table 1.

Simulated events for $qq \rightarrow VH$ plus zero or one jet production at next-to-leading order (NLO) were generated with the Powheg-Box v2 + GoSam + MiNLO generator [38, 41–43] (named Powheg+MiNLO in the rest of the article). The contribution from $gg \rightarrow ZH$ (gluon-induced) production was simulated using the leading-order (LO) Powheg-Box v2 matrix-element generator. An additional scale factor is applied to the $qq \rightarrow VH$ processes as a function of the vector boson’s transverse momentum to account for electroweak (EW) corrections at NLO. This makes use of the $VH$ differential cross-section computed with Hawk [72, 73]. The samples of simulated events include all final states where the Higgs boson decays into $b\bar{b}$ and the vector boson to a leptonic final state, including those with a $\tau$-lepton. The analysis has only a small acceptance for other Higgs boson production and decay modes which are therefore neglected. The mass of the Higgs boson was fixed at 125 GeV and the $H \rightarrow b\bar{b}$ branching fraction was fixed at 58%. The inclusive $pp \rightarrow VH$ cross-sections [44–50] were calculated at next-to-next-to-leading order (NNLO) (QCD) and
NLO (EW). Electroweak corrections include the photon-induced contributions, which are of the order of 5% for the $WH \rightarrow \ell\nu b\bar{b}$ process and 1% for the $ZH \rightarrow \ell\ell b\bar{b}$ process. For the gluon-induced $ZH$ production, the cross-section is calculated at next-to-leading order and next-to-leading-logarithm accuracy (NLO+NLL) in QCD [51–55]. This is then subtracted from the inclusive $pp \rightarrow ZH$ production cross-section to estimate the quark-induced contribution to the cross-section.

For the generation of $t\bar{t}$ at NLO, the Powheg-Box v2 generator [56] was used. Single top quark events in the $Wt$, $s$- and $t$-channels were generated using the Powheg-Box v1 generator [59, 65]. The top quark mass was set to 172.5 GeV. Events were filtered such that at least one $W$ boson in each event decays leptonically. The overall yield predicted for the $t\bar{t}$ process is rescaled according to the NNLO cross-section, including the resummation of soft gluon emission at next-to-next-to-leading-logarithm accuracy (NNLL) as available in Top++2.0 [58]. The overall yields predicted for single top quark production in the $s$, $t$, and $Wt$-channels are rescaled according to their respective NLO cross-sections [63, 64, 66].

Events containing $W$ or $Z$ bosons with jets ($V+$jets) were simulated using the Sherpa 2.2.1 generator. Matrix elements were calculated for up to two partons at NLO and four partons at LO using the OpenLoops [67] and Comix [68] matrix-element generators. The number of expected $V+$jets events is rescaled using the NNLO cross-sections [71].

Diboson $WZ$ and $ZZ$ (referred to as $VZ$) processes were generated using Sherpa 2.2.1, which calculates up to one additional parton at NLO and up to three additional partons at LO. The $WW$ process was generated using Sherpa 2.1.1, which calculates the inclusive production at NLO, and up to three additional partons at LO. The cross-sections from Sherpa at NLO are used to normalise the events.

Samples produced with alternative generators are used to estimate systematic uncertainties in the event modelling, as described in section 7.

4 Object and event selection

Events with two jets tagged as containing $b$-hadrons and with either zero, one or two charged leptons (electrons or muons) are selected in this analysis. In the following, the physics objects and the event selection for each channel are described.

4.1 Object reconstruction

Interaction vertices are reconstructed [74] from tracks measured by the inner detector. The vertex with the highest sum of squared transverse momenta of all associated tracks is selected as the primary vertex, whereas all others are considered to be pile-up vertices.

Electrons are reconstructed [75, 76] by applying a sliding-window algorithm to noise-suppressed clusters of energy deposited in the calorimeter and matching to a track in the inner detector. Their energy calibration is based primarily on a data sample of $Z \rightarrow e^+e^-$ events [77]. Electron candidates are required to satisfy criteria for the shower shape, track quality and track-to-cluster match, corresponding to either the loose or tight likelihood-based requirements, denoted “LooseLH” and “TightLH” in ref. [75]. All electrons are required to have $p_T > 7$ GeV and $|\eta| < 2.47$. Non-prompt and pile-up tracks are
rejected by requiring small transverse ($\text{IP}_{r \phi}$) and longitudinal ($\text{IP}_z$) impact parameters, defined with respect to the primary vertex position:\footnote{When computing impact parameters, the beam line is used to approximate the primary vertex position in the transverse plane.} tracks must have $|\text{IP}_{r \phi}|/\sigma_{\text{IP}_{r \phi}} < 5$ and $|\text{IP}_z| < 0.5$ mm, with $\sigma_{\text{IP}_{r \phi}}$ representing the uncertainty in the transverse impact parameter. A loose isolation requirement is applied: the electron track must be isolated from other tracks reconstructed in the inner detector, based on a variable cone size with $\Delta R_{\text{max}} = 0.2$, with a requirement that is tuned to yield a constant 99% efficiency as a function of electron $p_T$, corresponding to making the requirement looser with increasing values of $p_T$. Tight electrons are also required to pass a more stringent calorimeter-based isolation requirement, where the sum of the transverse energy of all the clusters of calorimeter cells, not associated with the electron candidate but found within a cone of $\Delta R = 0.2$ around the electron track, is required to be below 3.5 GeV.

Muons are reconstructed\cite{78, 79} as tracks in the inner detector matched to tracks in the muon spectrometer up to $|\eta| = 2.5$. Some acceptance is gained up to $|\eta| = 2.7$ using the muon spectrometer alone, and within the region $|\eta| < 0.1$ of limited muon-chamber acceptance, using tracks reconstructed in the inner detector that do not have a full matching track in the muon spectrometer, but have deposited energy in the calorimeter that is consistent with the energy loss of a muon. Two selection categories are defined: loose and medium, based on the respective muon identification criteria defined in ref.\cite{79}. All muon candidates are required to have $p_T > 7$ GeV, and not to be matched to an inner detector track that is likely to arise from a non-prompt muon or from pile-up, by applying impact parameter requirements similar to those for the electron selection: $|\text{IP}_{r \phi}|/\sigma_{\text{IP}_{r \phi}} < 3$ and $|\text{IP}_z| < 0.5$ mm. A loose isolation requirement is applied, based on the momenta of tracks in the inner detector which lie within a variable-size cone, with $\Delta R_{\text{max}} = 0.3$, around the muon track; analogously to the electron case, the requirement is tuned to yield a 99% efficiency for any value of $p_T$, corresponding to making the requirement looser with increasing values of $p_T$. For medium quality muons, a stringent track-based isolation requirement is applied, where the sum of the $p_T$ of all the tracks found within a cone of $\Delta R = 0.2$ around the muon track is required to be below 1.25 GeV.

Jets are reconstructed from noise-suppressed energy clusters in the calorimeter\cite{80} with the anti-$k_t$ algorithm\cite{81, 82} with radius parameter $R = 0.4$. The energies of the jets are calibrated using a jet energy scale correction (JES) derived from both simulation and in situ calibration using data\cite{83, 84}. Jet cleaning criteria are applied to find jets arising from non-collision sources or noise in the calorimeters and any event containing such a jet is removed\cite{85, 86}. Jets with $p_T$ below 60 GeV and with $|\eta| < 2.4$ have to pass a requirement on the jet vertex tagger (JVT)\cite{87}, a likelihood discriminant that uses track and vertex information in order to suppress jets originating from pile-up activity. Jets in the central region ($|\eta| < 2.5$) are required to have $p_T > 20$ GeV. For jets in the forward region ($2.5 \leq |\eta| < 4.5$), thus outside the acceptance of the inner detector, a stricter requirement of $p_T > 30$ GeV is applied in order to suppress jets from pile-up activity.
Jets in the central region can be tagged as containing $b$-hadrons by using a multivariate
discriminant (MV2c10) [88, 89] that combines information from an impact-parameter-based
algorithm, from the explicit reconstruction of a secondary vertex and from a multi-vertex
fitter that attempts to reconstruct the full $b$- to $c$-hadron decay chain. A significantly
improved algorithm, which also profits from the addition of the IBL detector, was developed
for Run 2 [89]. At the chosen working point, the improved algorithm provides nominal
light-flavour ($u,d,s$-quark and gluon) and $c$-jet misidentification efficiencies of 0.3% and
8.2%, respectively, for an average 70% $b$-jet tagging efficiency, as estimated from simulated
t$\bar{t}$ events for jets with $p_T > 20$ GeV and $|\eta| < 2.5$. The flavour tagging efficiencies in
simulation are corrected separately for $b$-, $c$- and light-flavour jets, based on the respective
data-based calibration analyses. The ratio of the efficiencies in data and simulation is close
to unity for $b$-jets, while more significant corrections are needed for $c$- and light-flavour
jets, up to $\approx 1.4$ and $\approx 2$, respectively.

Simulated jets are labelled according to which hadrons with $p_T > 5$ GeV are found
within a cone of size $\Delta R = 0.3$ around the jet axis. If a $b$-hadron is found the jet is labelled
as a $b$-jet. If no $b$-hadron is found, but a $c$-hadron is present, then the jet is labelled as a
c-jet. Otherwise the jet is labelled as a light (i.e., $u,d,s$-quark, or gluon) jet. Simulated
$V +$ jets events are categorised depending on the generator-level truth labels of the jets in
the event that are selected to form the Higgs boson candidate: $V + bb$, $V + bc$, $V + cc$, $V + bl$,
$V + cl$, $V + ll$ where $b$, $c$, $l$ stand for $b$-jet, $c$-jet and light-jet respectively. An inclusive
$V +$ heavy flavour ($V +$ HF) category is defined as containing the first four: $V + bb$, $V + bc$,
$V + cc$, $V + bl$. The $V + bb$ component is dominant: its fraction ranges from 70% to 90% of
$V +$ HF events, depending on the channel and analysis region.

Hadronically decaying $\tau$-leptons are reconstructed [90, 91] as jets from noise-suppressed
energy clusters, using the anti-$k_t$ algorithm with radius parameter $R = 0.4$. They are
required to have exactly one or three matching tracks in the inner detector within a cone
of size $\Delta R = 0.2$ around the jet axis, to have $p_T > 20$ GeV and $|\eta| < 2.5$, and to be outside
the transition region between the barrel and endcap calorimeters ($1.37 < |\eta| < 1.52$).
To reject jets being reconstructed and identified as $\tau$-leptons, a multivariate approach
using boosted decision trees is employed, based on information from the calorimeters and
from the tracking detectors; and the medium quality criteria described in ref. [91] are
applied. Hadronically decaying $\tau$-leptons are only used in the analysis in the overlap
removal procedure described at the end of this subsection. This has an impact on the
determination of the event’s jet multiplicity.

The uncertainty in the expected number of events depends on the size of the samples
of simulated events. The combination of processes with large production cross-section and
small selection efficiencies can make the production of samples exceeding the integrated
luminosity of the data challenging. For cases where the small selection efficiency is due to
the high rejection achieved by the application of $b$-tagging, a method called parameterised
tagging is applied. Unlike when explicitly applying the $b$-tagging algorithm (direct tag-
ging), in parameterised tagging all jets are kept but the event is weighted by the expected
probability for a jet with a certain flavour label ($b$, $c$ or light) to be tagged as a $b$-jet.
These probabilities are parameterised as a function of jet kinematics ($p_T$ and $\eta$) based on
a large sample of simulated $t\bar{t}$ events. Parameterised tagging is used for the $V + cc$, $V + c\bar{c}$, $V + ll$ and $WW$ samples, which simulate small background contributions ($< 2\%$ of the total background). For all other samples, direct tagging is applied.

In addition to the JES correction, two more corrections are applied to $b$-tagged jets. The muon-in-jet correction is applied when a medium quality muon with $p_T > 5$ GeV is found within $\Delta R = 0.4$ of a jet, to account for the presence of $b$- and $c$-hadron decays into muons which do not deposit their full energy in the calorimeter. Unlike in the lepton selection introduced previously, no isolation criteria are applied on such muons. When more than one muon is found, the one closest to the jet axis is chosen. The muon four-momentum is added to that of the jet, and the energy deposited by the muon in the calorimeter is removed. To further improve the jet response, a second correction, denoted $PtReco$, is applied as a function of jet $p_T$. This correction is derived separately for jets with or without a lepton (muon or electron) found within $\Delta R = 0.4$ of the jet axis, and it is based on the residual difference in jet response expected from the signal simulation between the reconstructed $b$-jets (with all corrections previously applied) and the corresponding truth jets (formed by clustering final-state particles taken from the Monte Carlo truth record, including muons and neutrinos). For jets without a matching lepton, the $PtReco$ correction increases the energy of jets with $p_T \sim 20$ GeV by 12% and the energy of those with $p_T > 100$ GeV by 1%, while a larger correction is observed for jets matched to a lepton, due to the missing neutrino energy.

In the 2-lepton channel, where the $ZH \rightarrow \ell\ell b\bar{b}$ event kinematics can be fully reconstructed, a per-event kinematic likelihood fit, the same which is described in detail in ref. [18] but with updated transfer functions, is used to improve the estimate of the energy of the two $b$-jets, in place of the $PtReco$ correction. These corrections result in an improved $m_{bb}$ mass distribution in the region of the Higgs boson signal, as illustrated in figure 1; the central value is moved closer to its nominal value, and the resolution is improved by up to about 40%.

The presence of neutrinos can be inferred by measuring the momentum imbalance in the event. This is measured by the missing transverse momentum $E_T^{miss}$, defined as the negative vector sum of the transverse momenta of electrons, muons and jets\(^3\) associated with the primary vertex. A soft term \(^{[93-95]}\) is added to include well-reconstructed tracks matched to the primary vertex that are not already matched to any of the physics objects.

The object reconstruction and identification algorithms do not always result in unambiguous identifications. An overlap removal procedure is therefore applied, with the following actions taken in sequence. Any hadronically decaying $\tau$-lepton reconstructed closer than $\Delta R = 0.2$ to an electron or muon is removed, except in cases where the muon is deemed to be of low quality. If a reconstructed muon shares an electron’s ID track, the electron is removed. Jets within a cone of size $\Delta R = 0.2$ around an electron are removed, since a jet is always expected from clustering an electron’s energy deposits in the calorimeter. Any electrons reconstructed within $\Delta R = \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{electron}})$ of the axis of any surviving jet are removed. Such electrons are likely to originate from

\(^3\)Hadronically decaying $\tau$-leptons are treated as jets in the measurement of the $E_T^{miss}$. 
Figure 1. Comparison of the $m_{bb}$ distributions as additional corrections are applied to the jet energy scale, shown for simulated events in the 2-lepton channel in the 2-jet and $p_T^Z > 150$ GeV region. A fit to a Bukin function [92] is superimposed on each distribution, and the resolution values and improvements are reported in the legend.

4.2 Event selection and categorisation

The online event selection relies on either the $E_{\text{miss}}^\tau$ or the single-charged-lepton triggers. Events passing the trigger selection and satisfying basic quality requirements are then categorised according to the charged lepton multiplicity, the vector boson’s transverse momentum, and jet multiplicity. Events are assigned to the 0-, 1- and 2-lepton channels depending on the number of charged leptons $\ell$, targeting the $ZH \to \nu\ell bb$, $WH \to \ell\nu bb$ and $ZH \to \ell\ell bb$ signatures, respectively. Although $\tau$-leptons from vector-boson decays are not targeted explicitly, they pass the selection with reduced efficiency through leptonic decays of the $\tau$-lepton into muons and electrons. All events are required to have at least two jets, and exactly two must pass the $b$-tagging requirement. The Higgs boson candidate is reconstructed from the two $b$-tagged jets and the highest-$p_T$ (leading) $b$-tagged jet is required to have $p_T > 45$ GeV.

The analysis covers the phase space at large Higgs boson (and equivalently vector boson) transverse momentum, which has the highest signal-to-background ratio. For the same reason, events are categorised according to the reconstructed vector boson’s transverse momentum $p_T^V$. This observable corresponds to $E_{\text{miss}}^\tau$ in the 0-lepton channel, to the size of
the vectorial sum of $E_T^{\text{miss}}$ and the charged lepton’s transverse momentum in the 1-lepton channel, and the transverse momentum of the 2-lepton system in the 2-lepton channel. In the 0- and 1-lepton channels a single region is defined, with $p_T^V > 150 \text{ GeV}$. In the 2-lepton channel two regions are considered, $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ and $p_T^V > 150 \text{ GeV}$.

Events are further split into two categories according to jet multiplicity. In the 0- and 1-lepton channels, events are considered with exactly two or exactly three jets. Events with four or more jets are rejected in these channels to reduce the large background arising from $t\bar{t}$ production. In the 2-lepton channel, extra sensitivity is gained by accepting events with higher jet multiplicity due to the lower level of the $t\bar{t}$ background, thus the categories become either exactly two jets or three or more jets. For simplicity, these two selection categories are referred to as the 2- and 3-jet categories for all three lepton channels.

The event selection criteria for the three channels are detailed below and summarised in table 2. The 1- and 2-lepton selections are both divided into two sub-channels depending on the flavour of the leptons: either electron or muon. There are small differences between these two sub-channels and these are mentioned when appropriate. The two sub-channels are merged to form the single 1- and 2-lepton channels used for the statistical analysis. The statistical analysis uses eight signal regions (SRs) and six control regions (CRs). Multivariate discriminants are used as the main observables to extract the signal, as described in section 5.

The predicted 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. The non-negligible acceptance for the $WH$ process in the 0-lepton channel is mostly due to events with hadronically decaying $\tau$-leptons produced in the $W$ decay, and the larger acceptance for the $gg \rightarrow ZH$ process with respect to $qq \rightarrow ZH$ is due to the harder $p_T^V$ distribution from the gluon-induced process.

4.2.1 Zero-lepton selection

The online event selection relies on an $E_T^{\text{miss}}$ trigger. The threshold for this trigger was 70 GeV for the 2015 data, and it was initially raised to 90 GeV and then to 110 GeV during 2016. In the offline analysis events are required to have no loose leptons and $E_T^{\text{miss}} > 150 \text{ GeV}$. When compared to the offline selection, the $E_T^{\text{miss}}$ trigger is fully efficient for $E_T^{\text{miss}} > 180 \text{ GeV}$, and it is $85 - 90\%$ efficient at $E_T^{\text{miss}} = 150 \text{ GeV}$, depending on the data taking period. The trigger efficiency is measured in $W + \text{jets}$ and $t\bar{t}$ events in data using an orthogonal set of single-muon triggers; these measurements are utilised to determine data-over-simulation scaling factors, used to correct the simulation. The scaling factors are within $5\%$ of unity and parameterised as a function of $E_T^{\text{miss}}$. A selection based on the scalar sum of the transverse momenta of the jets in the event, $H_T$, is used to remove a marginal region of phase space in which the trigger efficiency exhibits a small dependence on the jet multiplicity. For 2-jet events the requirement is $H_T > 120 \text{ GeV}$, and $H_T > 150 \text{ GeV}$ is required for 3-jet events.

In order to suppress the multi-jet background, which is mostly due to jets mismeasured in the calorimeters, four angular selection criteria are applied:
<table>
<thead>
<tr>
<th>Selection</th>
<th>0-lepton</th>
<th>1-lepton</th>
<th>2-lepton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td>$E_T^{miss}$ Single lepton</td>
<td>$E_T^{miss}$ Single lepton</td>
<td>$E_T^{miss}$ Single lepton</td>
</tr>
<tr>
<td>Leptons</td>
<td>0 loose leptons with $p_T &gt; 7$ GeV</td>
<td>1 tight electron $p_T &gt; 27$ GeV</td>
<td>2 loose leptons with $p_T &gt; 7$ GeV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 medium muon $p_T &gt; 25$ GeV</td>
<td>$\geq 1$ lepton with $p_T &gt; 27$ GeV</td>
</tr>
<tr>
<td>$E_T^{miss}$</td>
<td>$&gt; 150$ GeV</td>
<td>$&gt; 30$ GeV</td>
<td>$&gt; 30$ GeV</td>
</tr>
<tr>
<td>$m_{tt}$</td>
<td>--</td>
<td>--</td>
<td>$81$ GeV &lt; $m_{tt}$ &lt; $101$ GeV</td>
</tr>
<tr>
<td>Jets</td>
<td>Exactly 2 or 3 jets</td>
<td></td>
<td>Exactly 2 or 3 jets</td>
</tr>
<tr>
<td>Jet $p_T$</td>
<td></td>
<td>$&gt; 20$ GeV</td>
<td></td>
</tr>
<tr>
<td>$b$-jets</td>
<td></td>
<td>Exactly 2 $b$-tagged jets</td>
<td></td>
</tr>
<tr>
<td>Leading $b$-tagged jet $p_T$</td>
<td>$&gt; 45$ GeV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_T$</td>
<td>$&gt; 120$ (2 jets), $&gt; 150$ GeV (3 jets)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\min[\Delta \phi(E_T^{miss}, \text{jets})]$</td>
<td>$&gt; 20^\circ$ (2 jets), $&gt; 30^\circ$ (3 jets)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi(E_T^{miss}, bb)$</td>
<td>$&gt; 120^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi(b_1, b_2)$</td>
<td>$&lt; 140^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \phi(E_T^{miss}, E_T^{miss, \text{trk}})$</td>
<td>$&lt; 90^\circ$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_T^{\nu}$ regions</td>
<td></td>
<td>$&gt; 150$ GeV</td>
<td>$(75, 150]$ GeV, $&gt; 150$ GeV</td>
</tr>
<tr>
<td>Signal regions</td>
<td>$\checkmark$</td>
<td>$m_{bb} \geq 75$ GeV or $m_{top} \leq 225$ GeV</td>
<td>Same-flavour leptons</td>
</tr>
<tr>
<td>Control regions</td>
<td>--</td>
<td>$m_{bb} &lt; 75$ GeV and $m_{top} &gt; 225$ GeV</td>
<td>Different-flavour leptons</td>
</tr>
</tbody>
</table>

**Table 2.** Summary of the event selection in the 0-, 1- and 2-lepton channels.
Table 3. The cross-section times branching ratio (B) and acceptance for the three channels at \( \sqrt{s} = 13 \) TeV. The \( qq \)- and \( gg \)-initiated \( ZH \) processes are shown separately. The branching ratios are calculated considering only decays to muons and electrons for \( Z \to \ell \ell \), decays to all three lepton flavours for \( W \to \ell \nu \) and decays to all neutrino flavours for \( Z \to \nu \nu \). The acceptance is calculated as the fraction of events remaining in the combined signal and control regions after the full event selection.

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Process} & \text{Cross-section} \times \text{B} \text{ [fb]} & \text{Acceptance [%]} & \text{0-lepton} & \text{1-lepton} & \text{2-lepton} \\
\hline
qq \to ZH \to \ell \ell b \bar{b} & 29.9 & < 0.1 & < 0.1 & 7.0 \\
\hline
gg \to ZH \to \ell \ell b \bar{b} & 4.8 & < 0.1 & < 0.1 & 15.7 \\
\hline
qq \to WH \to \ell \nu b \bar{b} & 269.0 & 0.2 & 1.0 & - \\
\hline
qq \to ZH \to \nu \nu b \bar{b} & 89.1 & 1.9 & - & - \\
\hline
gg \to ZH \to \nu \nu b \bar{b} & 14.3 & 3.5 & - & - \\
\hline
\end{array}
\]

- \( \Delta \phi(E_{\text{T}}^{\text{miss}}, E_{\text{T,trl}}^{\text{miss}}) < 90^\circ \),
- \( \Delta \phi(b_1, b_2) < 140^\circ \),
- \( \Delta \phi(E_{\text{T}}^{\text{miss}}, bb) > 120^\circ \),
- \( \min[\Delta \phi(E_{\text{T}}^{\text{miss}}, \text{jets})] > 20^\circ \) for 2 jets, > 30° for 3 jets.

Here \( \Delta \phi(a, b) \) indicates the difference in azimuthal angle between objects \( a \) and \( b \); \( b_1 \) and \( b_2 \) are the two \( b \)-tagged jets forming the Higgs boson candidate’s dijet system \( bb \); \( E_{\text{T,trl}}^{\text{miss}} \) is defined as the missing transverse momentum calculated from the negative vector sum of the transverse momenta of tracks reconstructed in the inner detector and identified as originating from the primary vertex. The final selection is a requirement on the azimuthal angle between the \( E_{\text{T}}^{\text{miss}} \) vector and the closest jet.

4.2.2 One-lepton selection

For the electron sub-channel, events are selected using a logical OR of single-electron triggers with \( p_T \) thresholds of 24 GeV, 60 GeV and 120 GeV for the 2015 data and with increased thresholds of 26 GeV, 60 GeV and 140 GeV in 2016. The lowest-threshold trigger in 2016 includes isolation and identification requirements that are looser than any of the isolation and identification requirements applied in the analysis. These requirements are removed or relaxed for the higher-threshold triggers. The muon sub-channel uses the same \( E_{\text{T}}^{\text{miss}} \) triggers as the 0-lepton channel. Since muons are not included in the \( E_{\text{T}}^{\text{miss}} \) calculation at trigger level, in events where a muon is present this trigger is in effect selecting events based on \( p_T^V \), and is therefore fully efficient for values of \( p_T^V \) above 180 GeV. This trigger is preferred because it has an overall signal efficiency (with respect to the offline selection) of 98%, compared to \( \sim 80\% \) efficiency for the combination of single-muon triggers, which is due to the limited muon trigger chamber coverage in the central \( |\eta| \) region of the detector. Events are required to contain exactly one tight electron with \( p_T \) above 27 GeV (electron...
sub-channel) or one medium muon with $p_T$ above 25 GeV (muon sub-channel), and no additional loose leptons. In the electron sub-channel, where multi-jet production is a significant background, an additional selection of $E_T^{\text{miss}} > 30$ GeV is applied.

Control regions enhanced in the $W + \text{HF}$ background are defined for both the 2- and 3-jet categories. These are obtained by applying two additional selection requirements beyond the respective nominal selection criteria: $m_{lb} < 75$ GeV and $m_{\text{top}} > 225$ GeV. To calculate the reconstructed top quark mass, $m_{\text{top}}$, an estimate of the four-momentum of the neutrino from the $W$ boson decay is required. The vector $E_T^{\text{miss}}$ is assumed to give an estimate of the neutrino’s transverse momentum components and then $p_T^\nu$ can be determined up to a possible two-fold ambiguity by constraining the mass of the lepton-plus-neutrino system to be the $W$ boson mass.\footnote{In the case of negative discriminant in the quadratic equation, the $E_T^{\text{miss}}$ vector is shifted such that the discriminant becomes zero.} The top quark is then reconstructed by considering the reconstructed $W$ boson and one of the two $b$-tagged jets. The combination of $b$-tagged jet and $p_T^\nu$ minimising $m_{\text{top}}$ is selected. The requirement on the reconstructed top quark mass significantly reduces the contamination from $tt$ and single-top-quark events in the $W + \text{HF}$ CRs. The events in the control regions are removed from the corresponding signal regions. In the $W + \text{HF}$ CRs, between 75% and 78% of the events are expected to be from $W + \text{HF}$ production.

4.2.3 Two-lepton selection

Events are selected in the electron sub-channel using the same single-electron triggers as for the 1-lepton channel. For the muon sub-channel a logical OR of single-muon triggers with $p_T$ thresholds of 20 GeV and 40 GeV is used for 2015 data, and 24–26 GeV and 40–50 GeV for 2016 data, with the increase of the thresholds applied to cope with the increasing instantaneous luminosity. The lowest-threshold triggers include an isolation requirement that is removed for the higher-threshold triggers. The trigger efficiency with respect to the offline selection ranges from 97% to 99.5% for the electron sub-channel and from 87% to 90% for the muon sub-channel, depending on the $p_T$ region. To ensure that the trigger efficiency reached its plateau, the lepton that triggered the event is required to have $p_T > 27$ GeV.

Exactly two loose leptons of the same flavour are required. In dimuon events, the two muons are required to have opposite-sign charges. This is not used in the electron sub-channel, where the charge misidentification rate is not negligible. The invariant mass of the dilepton system must be consistent with the $Z$ boson mass: $81$ GeV $< m_{\ell\ell} < 101$ GeV. This requirement suppresses backgrounds with non-resonant lepton pairs, such as $tt$ and multi-jet production.

Control regions are defined to be very pure in $tt$ and $Wt$ background by applying the nominal selection but requiring an $e\mu$ lepton flavour combination instead of $ee$ or $\mu\mu$, and requiring the two leptons to have opposite-sign charges. The $t\bar{t}$ and $Wt$ events in these control regions are kinematically identical to those in the signal region, except for slight differences in acceptance between electrons and muons. These regions are called $e\mu$ CR in the following. In the $e\mu$ CRs, more than 99% of the events are expected to be from $tt$ and single top quark production, and between 88% and 97% from $t\bar{t}$ production alone.
### 4.3 Selection for the dijet-mass analysis

To validate the result of the multivariate analysis, a second analysis is performed where the multivariate discriminants are replaced by the dijet invariant mass of the two $b$-tagged jets, $m_{bb}$. This second analysis adopts the same objects and event selection criteria as described in table 2, with the additional selection criteria shown in table 4. With respect to the $p_T^V$ regions described earlier, the events with $p_T^V > 150$ GeV are further split into two categories: 150 GeV < $p_T^V$ ≤ 200 GeV and $p_T^V$ > 200 GeV. Events with $p_T^V$ ≤ 150 GeV are rejected if $\Delta R(b_1, b_2) > 3.0$, where $\Delta R(b_1, b_2)$ is the separation of the two $b$-tagged jets in the ($\eta, \phi$) plane. For 150 GeV < $p_T^V$ ≤ 200 GeV, the events are rejected if $\Delta R(b_1, b_2) > 1.8$. For $p_T^V > 200$ GeV events are rejected if $\Delta R(b_1, b_2) > 1.2$.

In the 1-lepton channel, since the low $m_{bb}$ range in the dijet mass spectrum provides sufficient information to constrain the $W + HF$ background normalisation, no dedicated $W + HF$ control region is defined. Also, a requirement on the $W$ boson’s transverse mass $m_W^T < 120$ GeV is used to suppress events from $t\bar{t}$ background. The $W$ boson’s transverse mass is defined as $m_W^T = \sqrt{2 p_T^V E_T^{miss} (1 - \cos(\Delta \phi(\ell, E_T^{miss})))}$, where $p_T^V$ is the lepton’s transverse momentum.

In the 2-lepton channel, the $t\bar{t}$ background is suppressed thanks to the additional requirement $E_T^{miss}/\sqrt{S_T} < 3.5\sqrt{\text{GeV}}$, where $S_T$ is defined as the scalar sum of the transverse momenta of all jets and leptons in the event. Events with $p_T^V > 150$ GeV in the $e\mu$ CR are used inclusively in $p_T^V$.

### 5 Multivariate analysis

Multivariate discriminants making use of boosted decision trees (BDTs) are constructed, trained and evaluated in each lepton channel and analysis region separately. Two versions of the BDTs, using the same input variables, are trained. The nominal version is designed to separate the $VH, H \to bb$ signal from the sum of the expected background processes, and is referred to as BDT$_{VH}$. A second one, which is used to validate the analysis, aims at
separating the $VZ, Z \rightarrow b\bar{b}$ diboson process from the sum of all other expected background processes (including $VH$), and is referred to as BDT$_{VZ}$.

The input variables used for the BDTs are chosen in order to maximise the separation in the $VH$ search. Starting from the dijet mass ($m_{bb}$), additional variables describing the event kinematics and topology are tried one at a time and the one yielding the best separation gain is added to the list. This procedure is repeated until adding more variables results in a negligible performance gain. The final selections of variables for the different channels are listed in table 5. The $b$-tagged jets are labelled in decreasing $p_T$ as $b_1$ and $b_2$, and $|\Delta\eta(b_1, b_2)|$ is their separation in pseudorapidity. In 3-jet events, the third jet is labelled as jet$_3$ and the mass of the 3-jet system is denoted $m_{bj3}$. The azimuthal angle between the vector boson and the system of $b$-tagged jets is denoted $\Delta\phi(V, bb)$, and their pseudorapidity separation is denoted $|\Delta\eta(V, bb)|$. In the 0-lepton channel, $m_{\text{eff}}$ is defined as the scalar sum of the transverse momenta of all jets and $E_T^{\text{miss}} (m_{\text{eff}} = H_T + E_T^{\text{miss}})$. In the 1-lepton channel, the angle between the lepton and the closest $b$-tagged jet in the transverse plane is denoted $\min[\Delta\phi(l, b)]$. In the 1-lepton channel, two variables are used to improve the rejection of the $t\bar{t}$ background: the rapidity difference between the $W$ and Higgs boson candidates, $|\Delta Y(V, bb)|$ and, assuming that the event is $t\bar{t}$, the reconstructed top quark mass, $m_{\text{top}}$. To construct the $|\Delta Y(V, bb)|$ variable, the four-vector of the neutrino in the $W$ boson decay is estimated as explained in section 4.2.2 for $m_{\text{top}}$. The distributions of input variables of the BDTs are compared between data and simulation, and good agreement is found within the uncertainties.

The Toolkit for Multivariate Data Analysis, TMVA [96], is used to train the BDTs, with values of the training parameters similar to those described in ref. [18]. 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. This procedure allows a reduction of the uncertainty due to the limited size of the MC samples by a factor of $\sqrt{2}$. A dedicated procedure is applied to transform the BDT output distributions to obtain a smoother distribution for the background processes and finer binning in the regions with the largest signal contribution, whilst ensuring that the statistical uncertainty of the simulated background is less than 20% in each bin. The binning procedure is described in more detail in ref. [18].

6 Estimation of the multi-jet background

The MC samples summarised in section 3 are used to model background processes with $W$ or $Z$ boson decays into leptons; these are defined as electroweak (EW) backgrounds in the following. Multi-jet backgrounds are produced with large cross-sections and thus, despite not providing genuine leptonic signatures, have the potential to contribute a non-negligible
Table 5. Variables used for the multivariate discriminant in each of the categories.

<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>$E_T^\text{miss}$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$E_T^\text{miss}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$p_T^{b_1}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$p_T^{b_2}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$m_{bb}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$\Delta R(b_1, b_2)$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \eta(b_1, b_2)</td>
<td>$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$\Delta \phi(V, bb)$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$m_{\text{eff}}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$\min[\Delta \phi(\ell, b)]$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$m_W$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$m_{\text{tt}}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$m_{t\bar{t}}$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$</td>
<td>\Delta Y(V, bb)</td>
<td>$</td>
<td>$\times$</td>
</tr>
</tbody>
</table>

Only in 3-jet events

6.1 0-lepton channel

As described in section 4, specific criteria are applied in the event selection to suppress the multi-jet backgrounds. A data-driven method is used to estimate the residual contribution. After removing the selection applied to the $\min[\Delta \phi(E_T^\text{miss, jets})]$ variable, a fit to this distribution in the 3-jet category is performed to extract the multi-jet contribution while allowing the $t\bar{t}$ and $Z + \text{jets}$ background normalisations to float. In multi-jet background events, a fake $E_T^\text{miss}$ can arise from a jet energy fluctuation, and it is expected that its direction is close to the direction of the poorly measured jet. Therefore, the $\min[\Delta \phi(E_T^\text{miss, jets})]$ variable is very effective in suppressing the multi-jet contribution, which is confined to low values of $x = \min[\Delta \phi(E_T^\text{miss, jets})]$ and is parameterised with a falling exponential $(\exp(-x/c))$. The parameter $c$ is determined in the fit itself, while the templates for the other backgrounds are taken directly from simulation. After the nominal selection criteria are applied, the residual multi-jet contamination within an $80 \text{ GeV} < m_{bb} < 160 \text{ GeV}$ mass window is found to be $\sim 10\%$ of the signal contribution and negligible ($< 0.1\%$) with respect to the total background. The BDT distribution for the multi-jet background is
estimated from the data at low \( \min[\Delta \phi(E_{T}^{\text{miss}}, \text{jets})] \), and found to have a shape similar to the one expected for the sum of the remaining backgrounds. The small multi-jet contribution is therefore absorbed in the floating normalisation factors of the EW backgrounds in the global likelihood fit. The same data-driven estimation technique cannot be used in the 2-jet region, where events at low values of \( \min[\Delta \phi(E_{T}^{\text{miss}}, \text{jets})] \) are removed by the other selection requirements. A multi-jet \textsc{Pythia8} MC sample generated with the A14 tune and NNPDF2.3LO PDFs is used to extrapolate the data-driven estimate from the 3- to the 2-jet region, with the extrapolation factor derived after removing any \( b \)-tagging requirement. The contribution in the 2-jet region is found to be negligible. Multi-jet production in the 0-lepton channel is therefore found to be a small enough background that it can be neglected in the global likelihood fit.

6.2 1-lepton channel

Both the electron and muon sub-channels have contributions from multi-jet events. The dominant contribution to this background stems from real muons or electrons from heavy-flavour hadrons that undergo semileptonic decays. In the electron sub-channel a second contribution arises from \( \gamma \rightarrow e^+e^- \) conversions of photons produced in the decay of neutral pions in jets, or directly from \( \pi^0 \) Dalitz decays. Although those leptons are not expected to be isolated, a small but non-negligible fraction passes the lepton isolation requirements. This background is estimated separately in the electron and muon sub-channels, and in the 2- and 3-jet categories, using similar procedures.

In each signal region, a template fit to the \( W \) boson candidate’s transverse mass \((m_{T}^{W})\) distribution is performed in order to extract the multi-jet yield. The variable \( m_{T}^{W} \) is chosen as it offers the clearest discrimination between the multi-jet and EW processes. The template used for the multi-jet contribution is obtained from data in a control region after subtraction of the residual EW contribution, based on MC predictions, while the template for the EW contribution in the signal region is obtained directly from MC predictions. The control region is enriched in multi-jet events that are kinematically close to the corresponding signal region but not overlapping with it, and is defined by applying the nominal selection but inverting the tight isolation requirement. Only one loose lepton is allowed to be present in the event, to keep orthogonality to the 2-lepton channel. To increase the statistical precision of the data-driven estimate, the number of required \( b \)-tags is reduced from two to one. The template fit applied in the signal region determines the normalisation of the multi-jet contribution, while the shape of the BDT discriminant (or of other relevant observables) is obtained analogously to the \( m_{T}^{W} \) template. Both the normalisation and shape derived for the BDT discriminant are then used in the global likelihood fit.

Since the efficiency of the tight isolation requirement on multi-jet events depends in general on lepton kinematics, and on the composition of the multi-jet background, the control regions that are based on inverting such a requirement provide biased estimators for the multi-jet templates in the corresponding signal regions. The templates are therefore corrected for such a bias, by applying event-by-event extrapolation factors that depend on lepton \( p_T \) and \( \eta \), and, in the electron sub-channel, also on the value of \( E_{T}^{\text{miss}} \). These extrapolation factors are derived in additional control regions where the 2- and 3-jet requirements
of the nominal selection are replaced by a 1-jet requirement, and the $b$-tagging requirement is removed. The extrapolation factors are computed as the ratio of the number of events with an isolated lepton to the number of events with a non-isolated lepton, after removing the MC-predicted EW background contribution.

The estimate of the normalisations of the $W + \text{jet}$ and top quark ($t\bar{t}$ and single top quark) background contributions in the signal region provided by Monte Carlo simulations is subject to significant uncertainties. In addition, the $m_W$ distributions of the $W + \text{jet}$ and top quark backgrounds are sufficiently different that a common normalisation factor induces a bias in the multi-jet estimate. The normalisation of these two backgrounds is therefore left free to be determined in the template fit used to extract the multi-jet contribution. In order to improve their relative separation, the fit to the $m_W$ distribution in the signal region is performed together with a fit to the overall yield in the corresponding $W + \text{HF}$ control region. Furthermore, in order to improve the statistical precision in the determination of the $W + \text{jet}$ and top quark background normalisation factors, the multi-jet template fit is performed simultaneously in the electron and muon sub-channels. This corresponds to performing separate fits for the two sub-channels, but with common $W + \text{jet}$ and top quark background normalisation factors. The multi-jet contribution in the 2-jet region is found to be 4.8% (4.6%) of the total background contribution in the electron (muon) sub-channel, while in the 3-jet region it is found to be 0.3% (0.5%). These estimates are subject to sizeable systematic uncertainties, which are described in section 7.

### 6.3 2-lepton channel

Requiring two isolated leptons with a dilepton invariant mass compatible with that of the $Z$ boson strongly suppresses the contributions from multi-jet events. The residual contribution is estimated using a fit to the dilepton mass distribution in a sample of events where the two lepton candidates have the same charge. The fit model includes expected contributions from EW backgrounds from simulation and an exponential model for the multi-jet background. An estimate is then made of the fraction of the background in a mass window around the $Z$ boson peak in the signal region that could be attributed to multi-jet events based on the assumption that the numbers of opposite-charge and same-charge events are equal for the multi-jet background. Inside a mass window $81 \text{GeV} < m_{\ell\ell} < 101 \text{GeV}$ the fraction of the background in the signal region coming from multi-jet events is estimated to be 0.03% and 0.2% for the muon and electron sub-channels, respectively. The residual multi-jet contamination within a $100 \text{GeV} < m_{bb} < 140 \text{GeV}$ mass window is found to be $\sim 8\%$ of the signal contribution, without an $m_{bb}$ resonant shape, and found to have a BDT shape similar to the one expected for the sum of the remaining backgrounds. The multi-jet contamination is also extracted in the $e\mu$ control region and found to be 0.3% of the total background. The multi-jet contribution in the 2-lepton channel is thus small enough to have a negligible impact on the signal extraction and is therefore not included in the global likelihood fit.
7 Systematic uncertainties

The sources of systematic uncertainty can be broadly divided into four groups: those of experimental nature, those related to the modelling of the simulated backgrounds, those related to the multi-jet background estimation, and those associated with the Higgs boson signal simulation. The finite size of the simulated background samples is also an important source of systematic uncertainty, and, whenever possible, generator-level filters are employed to enhance the amount of simulated events in the phase-space region that is most relevant for the analysis.

7.1 Experimental uncertainties

The dominant experimental uncertainties originate from the $b$-tagging simulation-to-data efficiency correction factors, from the jet energy scale corrections and the modelling of the jet energy resolution. The $b$-tagging simulation-to-data efficiency correction factors are derived [88] separately for $b$-jets, $c$-jets and light-flavour jets. All three correction factors depend on jet $p_T$ (or $p_T$ and $|\eta|$) and have uncertainties estimated from multiple sources. These are decomposed into uncorrelated components which are then treated independently, resulting in three uncertainties for $b$-jets and for $c$-jets, and five for light-flavour jets. The approximate size of the uncertainty in the tagging efficiency is 2% for $b$-jets, 10% for $c$-jets and 30% for light jets. Additional uncertainties are considered in the extrapolation of the $b$-jet efficiency calibration above $p_T = 300$ GeV and in the misidentification of hadronically decaying $\tau$-leptons as $b$-jets. The uncertainties in the jet energy scale and resolution are based on their respective measurements in data [84, 97]. The many sources of uncertainty in the jet energy scale correction are decomposed into 21 uncorrelated components which are treated as independent. An additional specific uncertainty is considered that affects the energy calibration of $b$- and $c$-jets.

Uncertainties in the reconstruction, identification, isolation and trigger efficiency of muons [79] and electrons [76], along with the uncertainty in their energy scale and resolution, are estimated based upon 13 TeV data. These are found to have only a small impact on the result. The uncertainties in the energy scale and resolution of the jets and leptons are propagated to the calculation of $E_T^{\text{miss}}$, which also has additional uncertainties from the scale, resolution and efficiency of the tracks used to define the soft term [94], along with the modelling of the underlying event. An uncertainty is assigned to the simulation-to-data $E_T^{\text{miss}}$ trigger scale factors to account for the statistical uncertainty in the measured scale factors and differences between the scale factors determined from $W + \text{jets}$ and $\bar{t}t$ events. The uncertainty in the luminosity is 2.1% for the 2015 data and 3.4% for the 2016 data, resulting in an uncertainty of 3.2% for the combined dataset. It is derived, following a methodology similar to that detailed in ref. [29], from a preliminary calibration of the luminosity scale using $x-y$ beam-separation scans performed in 2015 and 2016. The average number of interactions per bunch crossing is rescaled by 9% to improve the agreement between simulation with data, based on the measurement of the visible cross section in minimum-bias events [98], and an uncertainty, as large as the correction, is included.
7.2 Simulated background uncertainties

Modelling uncertainties are derived for the simulated backgrounds and broadly cover three areas: normalisation, acceptance differences that affect the relative normalisation between analysis regions with a common background normalisation, and the differential distributions of the most important kinematic variables. These uncertainties are derived either from particle-level comparisons between nominal and alternative samples using the RIVET [99] framework, or from comparisons to data in control regions. The particle-level comparisons are cross-checked with detector-level simulations whenever these are available, and good agreement is found. When acceptance uncertainties are estimated all the nominal and alternative samples are normalised using the same production cross-section. Such uncertainties are estimated by adding the differences between the nominal and alternative samples in quadrature. Shape uncertainties are considered in each of the analysis regions separately, with the samples scaled to have the same normalisation in each region. In this case, the uncertainty is taken from the alternative generator which has the largest shape difference compared to the nominal sample. Shape uncertainties are only derived for the $m_{bb}$ and $p_{T}^{V}$ variables, as it was found that it is sufficient to only consider the changes induced in these variables by an alternative generator to cover the overall shape variation of the $BDT_{VH}$ discriminant. The systematic uncertainties affecting the modelling of the background samples are reported in tables 6 and 7, and the specific details of how the uncertainties are estimated are provided below for each simulated background sample.

**$V + \text{jets}$ production.** The $V + \text{jets}$ backgrounds are subdivided into three different components based upon the jet flavour labels of the two $b$-tagged jets in the event. The main background contributions ($V + bb$, $V + bc$, $V + bl$ and $V + cc$) are jointly considered as the $V + \text{HF}$ background. Their overall normalisation, separately in the 2- and 3-jet categories, is free to float in the global likelihood fit, as detailed in section 8. The remaining flavour components, $V + cl$ and $V + ll$, make up less than ~1% of the background in each analysis region, so only uncertainties in the normalisation of these backgrounds are included.

Acceptance uncertainties are estimated for the relative normalisations of the different regions that share a common floating normalisation parameter. In the case of the $W + \text{HF}$ background, this includes the uncertainties in the ratio of the event yield in the 0-lepton channel to that in the 1-lepton channel and, in the 1-lepton channel, in the ratio of the event yield in the $W + \text{HF}$ control region to that in the signal region. For the $Z + \text{HF}$ background, there is an uncertainty in the ratio of the event yield in the 0-lepton channel to that in the 2-lepton channel. These ratio uncertainties act as effective extrapolation uncertainties from one region to another.

Uncertainties are also estimated in the relative normalisation of the four heavy-flavour components that make up the $V + \text{HF}$ background. These are taken as uncertainties in the $bc$, $cc$ and $bl$ yields compared to the dominant $bb$ yield and are estimated separately for the 0- and 1-lepton channels in the case of $W + \text{HF}$ and separately for the 0-lepton, 2-lepton 2-jet and 2-lepton 3-jet regions in the case of $Z + \text{HF}$.

The normalisation and acceptance uncertainties are all calculated by adding the differences between the nominal SHERPA 2.2.1 sample and its associated systematic varia-
<table>
<thead>
<tr>
<th>Source</th>
<th>Normalisation</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Z + jets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z + ll normalisation</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>Z + cl normalisation</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>Z + bb normalisation</td>
<td>Floating (2-jet, 3-jet)</td>
<td></td>
</tr>
<tr>
<td>Z + bc-to-Z + bb ratio</td>
<td>30 – 40%</td>
<td></td>
</tr>
<tr>
<td>Z + cc-to-Z + bb ratio</td>
<td>13 – 15%</td>
<td></td>
</tr>
<tr>
<td>Z + bl-to-Z + bb ratio</td>
<td>20 – 25%</td>
<td></td>
</tr>
<tr>
<td>0-to-2 lepton ratio</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>m_{bb}, p_T^V</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td><strong>W + jets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W + ll normalisation</td>
<td>32%</td>
<td></td>
</tr>
<tr>
<td>W + cl normalisation</td>
<td>37%</td>
<td></td>
</tr>
<tr>
<td>W + bb normalisation</td>
<td>Floating (2-jet, 3-jet)</td>
<td></td>
</tr>
<tr>
<td>W + bl-to-W + bb ratio</td>
<td>26% (0-lepton) and 23% (1-lepton)</td>
<td></td>
</tr>
<tr>
<td>W + bc-to-W + bb ratio</td>
<td>15% (0-lepton) and 30% (1-lepton)</td>
<td></td>
</tr>
<tr>
<td>W + cc-to-W + bb ratio</td>
<td>10% (0-lepton) and 30% (1-lepton)</td>
<td></td>
</tr>
<tr>
<td>0-to-1 lepton ratio</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>W + HF CR to SR ratio</td>
<td>10% (1-lepton)</td>
<td></td>
</tr>
<tr>
<td>m_{bb}, p_T^V</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td><strong>t\bar{t}</strong> (all are uncorrelated between the 0+1 and 2-lepton channels)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t\bar{t} normalisation</td>
<td>Floating (0+1 lepton, 2-lepton 2-jet, 2-lepton 3-jet)</td>
<td></td>
</tr>
<tr>
<td>0-to-1 lepton ratio</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>2-to-3-jet ratio</td>
<td>9% (0+1 lepton only)</td>
<td></td>
</tr>
<tr>
<td>W + HF CR to SR ratio</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>m_{bb}, p_T^V</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td><strong>Single top quark</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-section</td>
<td>4.6% (s-channel), 4.4% (t-channel), 6.2% (Wt)</td>
<td></td>
</tr>
<tr>
<td>Acceptance 2-jet</td>
<td>17% (t-channel), 35% (Wt)</td>
<td></td>
</tr>
<tr>
<td>Acceptance 3-jet</td>
<td>20% (t-channel), 41% (Wt)</td>
<td></td>
</tr>
<tr>
<td>m_{bb}, p_T^V</td>
<td>S (t-channel, Wt)</td>
<td></td>
</tr>
<tr>
<td><strong>Multi-jet (1-lepton)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalisation</td>
<td>60 – 100% (2-jet), 100 – 400% (3-jet)</td>
<td></td>
</tr>
<tr>
<td>BDT template</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Summary of the systematic uncertainties in the background modelling for Z + jets, W + jets, t\bar{t}, single top quark and multi-jet production. An “S” symbol is used when only a shape uncertainty is assessed. The regions for which the normalisations float independently are listed in brackets.
<table>
<thead>
<tr>
<th>ZZ</th>
<th>Normalisation</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-to-2 lepton ratio</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>Acceptance from scale variations (var.)</td>
<td>10 – 18% (Stewart-Tackmann jet binning method)</td>
</tr>
<tr>
<td></td>
<td>Acceptance from PS/UE var. for 2 or more jets</td>
<td>5.6% (0-lepton), 5.8% (2-lepton)</td>
</tr>
<tr>
<td></td>
<td>Acceptance from PS/UE var. for 3 jets</td>
<td>7.3% (0-lepton), 3.1% (2-lepton)</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, $p_T^V$, from scale var.</td>
<td>S (correlated with WZ uncertainties)</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, $p_T^V$, from PS/UE var.</td>
<td>S (correlated with WZ uncertainties)</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, from matrix-element var.</td>
<td>S (correlated with WZ uncertainties)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WZ</th>
<th>Normalisation</th>
<th>26%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-to-1 lepton ratio</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>Acceptance from scale var.</td>
<td>13 – 21% (Stewart-Tackmann jet binning method)</td>
</tr>
<tr>
<td></td>
<td>Acceptance from PS/UE var. for 2 or more jets</td>
<td>3.9%</td>
</tr>
<tr>
<td></td>
<td>Acceptance from PS/UE var. for 3 jets</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, $p_T^V$, from scale var.</td>
<td>S (correlated with ZZ uncertainties)</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, $p_T^V$, from PS/UE var.</td>
<td>S (correlated with ZZ uncertainties)</td>
</tr>
<tr>
<td></td>
<td>$m_{bb}$, from matrix-element var.</td>
<td>S (correlated with ZZ uncertainties)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WW</th>
<th>Normalisation</th>
<th>25%</th>
</tr>
</thead>
</table>

Table 7. Summary of the systematic uncertainties in the background modelling for diboson production. “PS/UE” indicates parton shower / underlying event. An “S” symbol is used when only a shape uncertainty is assessed. When determining the ($W/Z$)Z diboson production signal strength, the normalisation uncertainties in ZZ and WZ production are removed.

tions in quadrature, including a variation of (i) the renormalisation scale by factors of 0.5 and 2; (ii) the factorisation scale by factors of 0.5 and 2; (iii) the CKKW merging scale from 30 GeV to 15 GeV; (iv) the parton-shower/resummation scale by factors of 0.5 and 2. In addition, the difference between the SHERPA 2.2.1 nominal sample and an alternative sample produced with a different matrix-element generator is added in quadrature to the rest to yield the total uncertainty. The alternative sample is produced with MADGRAPH5_aMC@NLO v2.2.2 [100], with up to four extra partons at LO, and interfaced to PYTHIA 8.212; the A14 tune is used together with the NNPDF2.3LO PDF set.

Uncertainties in the shapes of the $m_{bb}$ and $p_T^V$ distributions are estimated for $Z + \text{HF}$ by comparing the $Z + \text{jets}$ background to data in signal-depleted regions with a very high $Z + \text{jets}$ purity, specifically the 1- and 2-tag regions of the 2-lepton channel, with the $m_{bb}$ region around the Higgs boson mass excluded in the 2-tag case. In order to remove most of the residual $t\bar{t}$ contamination, a selection requirement is made on $E_T^{\text{miss}}/\sqrt{S_T} < 3.5 \text{ GeV}$ as done for the dijet-mass analysis.

For the $W + \text{HF}$ background, due to the limited number of events in the dedicated control region, shape uncertainties are based on the same systematic uncertainty sources as for the normalisation and acceptance uncertainties; in all event categories, since scale variations are found to have a minor effect on the shapes of the distributions, the systematic
uncertainties are dominated by the comparison of the nominal Sherpa 2.2.1 sample with Madgraph5_aMC@NLO v2.2.2.

**$t\bar{t}$ production.** Uncertainties are derived from comparing the nominal sample (POWHEG+PYTHIA8) to alternative samples with different parton-shower generation (POWHEG+HERWIG7 with the H7-UE-MMHT tune [101, 102]), matrix-element generation (MADGRAPH5_aMC@NLO+PYTHIA8) and settings of the nominal generator designed to increase or decrease the amount of radiation. Due to the significantly different regions of phase space probed, the $t\bar{t}$ background in the 0- and 1-lepton channels (jointly referred to as 0+1 lepton in the following) is considered independently from the $t\bar{t}$ background in the 2-lepton channel; different overall floating normalisation factors are considered, and acceptance uncertainties are derived separately and taken as uncorrelated between the 0+1 and 2-lepton channels. For the 0+1 lepton channels, uncertainties are considered in the normalisation ratios of the 3-jet and 2-jet regions, of the $W+HF$ control region and signal region, and of the 1-lepton and 0-lepton channels. These uncertainties are estimated by comparing the difference between the ratios of the yields in the two regions under consideration in the alternative $t\bar{t}$ samples and those measured in the nominal sample. The differences between the nominal and each of the alternative samples are summed in quadrature to provide an overall uncertainty. For the 2-lepton channel, the normalisations in the 2- and 3-jet regions are both left floating, and are effectively determined in their respective $e\mu$ control regions. Uncertainties in the shapes of the $p_T^V$ and $m_{bb}$ distributions are estimated in the 0+1 and 2-lepton channels separately. The difference between the nominal sample and MADGRAPH5_aMC@NLO provides by far the largest variation, and is therefore considered as a systematic uncertainty in the shapes of these distributions.

**Single top quark production.** In the $Wt$ and $t$-channels, uncertainties are derived in the normalisation, acceptance and shapes of the $m_{bb}$ and $p_T^V$ distributions. The $s$-channel only has a normalisation uncertainty derived as its contribution is negligible overall.

For the $t$-channel, the nominal samples (POWHEG+PYTHIA6) are compared to alternative samples, which are similar to those used in the $t\bar{t}$ case using different parton-shower generation (POWHEG+HERWIG++ with the CTEQ6L1-UE-EE-5 tune), and matrix-element generation (MADGRAPH5_aMC@NLO+HERWIG++ with the CTEQ6L1-UE-EE-5 tune). For the $Wt$ channel, uncertainties related to the interference between the $Wt$ and $t\bar{t}$ production processes are assessed by using a diagram subtraction scheme instead of the nominal diagram removal scheme [65, 103]. For both the $t$- and $Wt$-channels, the settings of the nominal generator are varied so as to maximise or minimise the amount of radiation. The normalisation uncertainties take into account variations of the renormalisation and factorisation scales, $\alpha_S$ and PDFs. Uncertainties in the acceptance in both the 2- and 3-jet regions are derived by comparing the alternative generators and summing the differences with respect to the nominal sample in quadrature. Shape uncertainties are derived for the $m_{bb}$ and $p_T^V$ distributions. These uncertainties cover all the differences in the shapes of the kinematic distributions investigated by comparing nominal and alternative samples.
Diboson production. The diboson backgrounds are composed of three distinct processes, $WZ$, $WW$ and $ZZ$ production. Given the small contribution from $WW$ production (< 0.1% of the total background) only a normalisation uncertainty is assigned. The more important contributions from the $WZ$ and $ZZ$ backgrounds have uncertainties derived for the overall normalisation, the relative acceptance between regions and for the $m_{bb}$ and $p_T^V$ shapes. Uncertainties are derived by comparing the nominal sample (SHERPA 2.2.1) to the alternative samples with varied factorisation, renormalisation and resummation scales, and using the Stewart-Tackmann method [104] to estimate scale variation uncertainties for the acceptance in the jet multiplicity categories. Additional uncertainties in the overall acceptance, in the relative acceptance across jet multiplicities and in the shape of the $m_{bb}$ and $p_T^V$ distributions are estimated in the parton-shower and underlying-event model. These are estimated by considering the difference between POWHEG+PYTHIA8 and POWHEG+HERWIG++ with the CTEQ6L1-UE-EE-5 tune, as well as changes in the PYTHIA8 parton-shower tune. The envelope of the two effects is considered to define these uncertainties. A systematic uncertainty in the shape of the $m_{bb}$ distribution results from the comparison of SHERPA 2.2.1 and POWHEG+PYTHIA8. This changes the shape of the $m_{bb}$ distribution for values in the range $100 - 130$ GeV by $10 - 20\%$. Acceptance uncertainties are derived for the ratio of 0-to-1 lepton channels and for the ratio of the 2-to-3 jet regions for $WZ$ production. In the $ZZ$ production case the acceptance uncertainties are derived for the ratio of the 0-to-2 lepton channels and of the 2-to-3 jet regions. Uncertainties in the acceptance and $m_{bb}$ or $p_T^V$ shapes of the diboson background due to PDF and $\alpha_S$ variations were found to have a negligible impact.

7.3 Multi-jet background uncertainties

The multi-jet background in the 1-lepton channel is estimated from data as outlined in section 6. Systematic uncertainties can have an impact on the multi-jet estimates in two ways: either changing the $m_{W_T}^W$ distributions used in the multi-jet template fits, therefore impacting the extracted multi-jet normalisations, or directly changing the multi-jet BDT distributions used in the global likelihood fit. Several sources of uncertainty are considered, uncorrelated between the electron and muon sub-channels. The respective variations are added in quadrature for the normalisations, or considered as separate shape uncertainties. The variations are obtained by changing the definition of the multi-jet control region (2 versus 1 $b$-tag, more stringent isolation requirements, a different single-electron trigger to probe a potential trigger bias in the isolation requirements); removing the bias correction that makes use of the $p_{T^-}$, $\eta$, and $E_{T}^{\text{miss}}$- dependent extrapolation factors derived in the (1-jet, 0 $b$-tag) region; varying the normalisation of the contamination from the top ($t\bar{t}$ and $Wt$) and $V$+jets processes in the multi-jet control region. In addition, the following sources of uncertainty that only have an impact on the multi-jet normalisation, are considered: use of the $E_{T}^{\text{miss}}$ variable instead of $m_{W_T}^W$ for the multi-jet template fit and, for the electron sub-channel only, the inclusion of the $E_{T}^{\text{miss}} < 30$ GeV region, which significantly enhances the multi-jet contribution in the template fit.
### 7.4 Signal uncertainties

The signal samples are normalised using their inclusive cross-sections, as described in section 3, and an additional scale factor computed using the HAWK generator is applied as a function of $p_T$ to correct for the sizeable impact of the NLO (EW) corrections to the $p_T$ distributions. The systematic uncertainties that affect the modelling of the signal are summarised in table 8.

Uncertainties in the calculations of the $VH$ production cross-sections and the $H \rightarrow bb$ branching ratio are assigned following the recommendations of the LHC Higgs Cross section working group [15, 54, 55, 105, 106]. The uncertainties in the overall $VH$ production cross-section from missing higher-order terms in the QCD perturbative expansion are obtained by varying the renormalisation scale $\mu_R$ and factorisation scale $\mu_F$ independently, from 1/3 to 3 times their original value. The PDF+$\alpha_S$ uncertainty in the overall $VH$ production cross-section is calculated from the 68% CL interval using the PDF4LHC15_nnlo_mc PDF set. The latest recommendations of the LHC Higgs working group [107] do not distinguish between uncertainties in $qq \rightarrow ZH$ production and $gg \rightarrow ZH$ production. To obtain the scale uncertainties separately for these two processes, it is assumed that the uncertainty in $qq \rightarrow ZH$ production is identical to the uncertainty in $WH$ production. The $gg \rightarrow ZH$ production uncertainty is then derived such that the sum in quadrature of the $qq \rightarrow ZH$ and $gg \rightarrow ZH$ production uncertainties (considering their respective production cross-sections) amount to the scale uncertainty in the overall $ZH$ production. Since the PDF+$\alpha_S$ uncertainty is larger for $WH$ production than $ZH$ production, the method used for the scale uncertainty cannot be used for this uncertainty. The PDF+$\alpha_S$ uncertainty in the $gg \rightarrow ZH$ production is taken from previous recommendations [15] and the uncertainty in the $qq \rightarrow ZH$ production is taken from the latest recommendation [107].

Another systematic uncertainty in the overall $VH$ cross-section originates from missing higher-order electroweak corrections. This is estimated as the maximum variation among
three quantities: the maximum size expected for the missing NNLO EW effects (1%), the
size of the NLO EW correction and the uncertainty in the photon-induced cross-section
relative to the total (W/Z)H cross-section described in section 3. The systematic uncer-
tainty in the $H \rightarrow b\bar{b}$ branching ratio is 1.7% [13]. This uncertainty takes into account
missing higher-order QCD and EW corrections as well as uncertainties in the $b$-quark mass
and in the value of $\alpha_S$.

Acceptance and shape systematic uncertainties are derived to account for missing
higher-order QCD and EW corrections, for PDF+$\alpha_S$ uncertainty, and for variations of
the parton-shower and underlying-event models. Uncertainties in the acceptance and in
the shape of the $m_{bb}$ and $p_T^V$ distributions, originating from missing higher-order terms in
QCD, are estimated by comparing the nominal samples to those generated with weights
corresponding to varied factorisation and renormalisation scales applied. The Stewart-
Tackmann method is used to assign scale variation uncertainties in the acceptance in the
jet multiplicity categories. Uncertainties due to the parton-shower and underlying-event
models are estimated by considering the difference between POWHEG+MiNLO+Pythia8
and POWHEG+MiNLO+Herwig7 with the H7-UE-MMHT tune, as well as changes in
the Pythia8 parton-shower tune. The latter effect is assessed in events generated with
Madgraph5_aMC@NLO and showered with Pythia8, using the A14 tune and its
variations. The envelope of the two effects is considered to define uncertainties separately in the
overall acceptance, in the relative acceptance across jet multiplicities and in the shape of the
$m_{bb}$ and $p_T^V$ distributions. The PDF+$\alpha_S$ uncertainty in the acceptance between regions
and in the $m_{bb}$ and $p_T^V$ shapes is estimated applying the PDF4LHC15_30 PDF set and its
uncertainties, according to the PDF4LHC recommendations [37].

8 Statistical analysis

8.1 Analysis of the 13 TeV data

A statistical fitting procedure based on the Roostats framework [108, 109] is used to extract
the strength of the Higgs boson signal 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 uncer-
tainties.

The different regions entering the likelihood fit are summarised in table 9. The primary
inputs to the global fit are the BDT$_{VH}$ discriminants in the eight 2-b-tag signal regions
defined by the three lepton channels, up to two $p_T^V$ intervals and the two jet multiplicity
categories. Additional inputs are the event yields in the two $W + HF$ control regions in the
1-lepton channel subdivided into the two number-of-jet categories, and the $m_{bb}$ distributions
or the event yields for the four $e\mu$ control regions defined by the two $p_T^V$ intervals and the
two number-of-jet categories. The electron and muon sub-channels are summed up and
combined before entering the fit. Altogether, there are 141 bins in the 14 regions used
in the global fit. In addition to the global fit with all channels combined, separate 0-, 1-
Table 9. The distributions used in the global likelihood fit for the signal regions (SR) and control regions (CR) for all the categories in each channel, for the nominal multivariate analysis.

<table>
<thead>
<tr>
<th>Channel</th>
<th>SR/CR</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>75 GeV &lt; $p_T^V$ &lt; 150 GeV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 jets</td>
</tr>
<tr>
<td>0-lepton</td>
<td>SR</td>
<td>—</td>
</tr>
<tr>
<td>1-lepton</td>
<td>SR</td>
<td>—</td>
</tr>
<tr>
<td>2-lepton</td>
<td>SR</td>
<td>BDT</td>
</tr>
<tr>
<td>1-lepton</td>
<td>W + HF CR</td>
<td>—</td>
</tr>
<tr>
<td>2-lepton</td>
<td>$\ell\mu$ CR</td>
<td>$m_{bb}$</td>
</tr>
</tbody>
</table>

and 2-lepton channel fits are performed, where only the analysis regions specific to a single channel are considered and a channel-specific signal strength is obtained.

The effect of systematic uncertainties in the signal and background predictions 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 $\mu$ and $\theta$. For each NP, the prior is added as a penalty term to the likelihood, $\mathcal{L}(\mu, \theta)$, which decreases as soon as the nuisance parameter $\theta$ is shifted away from its nominal value. The statistical uncertainties of background predictions from simulation are included through one nuisance parameter per bin, using the Beeston-Barlow technique [110].

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

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

where $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximise the likelihood, and $\hat{\theta}_\mu$ are the nuisance parameter values that maximise the likelihood for a given $\mu$. 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 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 $q_\mu$ by one unit. Several tests have been carried out to verify that the likelihood function provides an unbiased estimator of the signal strength and of its associated uncertainty. Expected results are obtained in the same way as the observed results by replacing the data in each input bin by the prediction from simulation with all NPs set to their best-fit values, as obtained from the fit to the data, except for the signal strength parameter, which is kept at its nominal value.

The data have sufficient statistical power to constrain the largest background normalisation NPs, which are left free to be determined in the fit without having priors. This applies to the $t\bar{t}$, $W + \text{HF}$ and $Z + \text{HF}$ processes. The corresponding normalisation factors
expressed with respect to their expected nominal value and resulting from the global fit to the 13 TeV data, are shown in table 10. As stated in section 7, the $t\bar{t}$ background is normalised independently for the 2-lepton channel and for the 0- and 1-lepton channels. In the 2-lepton channel, the $t\bar{t}$ background is almost entirely due to events in which both top quarks decay into $(W \rightarrow \ell \nu)b$ (dileptonic decays) with all final-state objects detected (apart from the neutrinos). In the 0- and 1-lepton channels, it is in part due to dileptonic decays with one or two of the leptons (often a $\tau$-lepton) undetected, and in part due to cases where one of the top quarks decays into $(W \rightarrow q\bar{q})b$ (semileptonic decays) with at least one undetected light- or $c$-quark jet. Furthermore, the $p_T^{V}$ range probed is different in the 0- and 1-lepton channels: $p_T^{V} > 150$ GeV in contrast to $p_T^{V} > 75$ GeV in the 2-lepton channel. For the $Z + HF$ and $W + HF$ backgrounds, the data have enough statistical power to constrain the normalisations in the 2-jet and in the 3-jet categories independently. The normalisation factors for these backgrounds can deviate significantly from one due to the large theoretical uncertainty in the cross-sections of the contributing processes.

The systematic uncertainties are encoded in variations of the nominal BDT$_{VH}$ or $m_{bb}$ templates, and of the nominal yields across analysis categories, 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. To reduce the complexity of the fit, systematic uncertainties that have a negligible impact on the final results are pruned away, region by region. Studies were performed to verify that the smoothing and pruning procedures do not induce any bias in the result. More details about the smoothing and pruning procedures can be found in ref. [18].

<table>
<thead>
<tr>
<th>Process</th>
<th>Normalisation factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$ 0- and 1-lepton</td>
<td>$0.90 \pm 0.08$</td>
</tr>
<tr>
<td>$t\bar{t}$ 2-lepton 2-jet</td>
<td>$0.97 \pm 0.09$</td>
</tr>
<tr>
<td>$t\bar{t}$ 2-lepton 3-jet</td>
<td>$1.04 \pm 0.06$</td>
</tr>
<tr>
<td>$W + HF$ 2-jet</td>
<td>$1.22 \pm 0.14$</td>
</tr>
<tr>
<td>$W + HF$ 3-jet</td>
<td>$1.27 \pm 0.14$</td>
</tr>
<tr>
<td>$Z + HF$ 2-jet</td>
<td>$1.30 \pm 0.10$</td>
</tr>
<tr>
<td>$Z + HF$ 3-jet</td>
<td>$1.22 \pm 0.09$</td>
</tr>
</tbody>
</table>

Table 10. Factors applied to the nominal normalisations of the $t\bar{t}$, $W + HF$ and $Z + HF$ backgrounds, as obtained from the global fit to the 13 TeV data for the nominal multivariate analysis, used to extract the Higgs boson signal. The errors include the statistical and systematic uncertainties.
### Table 11.
Breakdown of the contributions to the uncertainties in $\hat{\mu}$. The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. The $b$-tagging extrapolation uncertainty refers to the extrapolation of the $b$-jet calibration above $p_T = 300$ GeV.

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>$\sigma_{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.39</td>
</tr>
<tr>
<td>Statistical</td>
<td>0.24</td>
</tr>
<tr>
<td>Systematic</td>
<td>0.31</td>
</tr>
</tbody>
</table>

#### Experimental uncertainties

<table>
<thead>
<tr>
<th>Category</th>
<th>$\sigma_{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jets</td>
<td>0.03</td>
</tr>
<tr>
<td>$E_T^{\text{miss}}$</td>
<td>0.03</td>
</tr>
<tr>
<td>Leptons</td>
<td>0.01</td>
</tr>
<tr>
<td>$b$-tagging</td>
<td></td>
</tr>
<tr>
<td>$b$-jets</td>
<td>0.09</td>
</tr>
<tr>
<td>$c$-jets</td>
<td>0.04</td>
</tr>
<tr>
<td>light jets</td>
<td>0.04</td>
</tr>
<tr>
<td>extrapolation</td>
<td>0.01</td>
</tr>
<tr>
<td>Pile-up</td>
<td>0.01</td>
</tr>
<tr>
<td>Luminosity</td>
<td>0.04</td>
</tr>
</tbody>
</table>

#### Theoretical and modelling uncertainties

<table>
<thead>
<tr>
<th>Category</th>
<th>$\sigma_{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>0.17</td>
</tr>
<tr>
<td>Floating normalisations</td>
<td>0.07</td>
</tr>
<tr>
<td>$Z$ + jets</td>
<td>0.07</td>
</tr>
<tr>
<td>$W$ + jets</td>
<td>0.07</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>0.07</td>
</tr>
<tr>
<td>Single top quark</td>
<td>0.08</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.02</td>
</tr>
<tr>
<td>Multijet</td>
<td>0.02</td>
</tr>
<tr>
<td>MC statistical</td>
<td>0.13</td>
</tr>
</tbody>
</table>

In order to understand the effect of systematic uncertainties on the final results, the breakdown of the contributions to the uncertainties in $\hat{\mu}$ is reported in table 11. The individual sources of systematic uncertainty detailed in section 7 are combined into categories. To assess the contribution of a category to the total systematic uncertainty, all NPs associated with the uncertainties within the category are fixed to their fitted values and the fit is repeated. The difference in quadrature between the uncertainties for $\hat{\mu}$ from this fit and from the nominal fit provides an estimate of the systematic uncertainty attached to the considered category of uncertainties. As shown in the table, the systematic uncertainties for the modelling of the signal play a dominant role, followed by the uncertainty due to the limited size of the simulated samples, the modelling of the backgrounds and the $b$-jet tagging uncertainty.
Table 12. The distributions used in the global likelihood fit for the dijet-mass analysis, for the signal regions (SR) and control regions (CR), for all the categories in each channel. The two regions marked with * (†) are merged into a single region, to reduce statistical uncertainties.

<table>
<thead>
<tr>
<th>Channel</th>
<th>SR/CR</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>75 GeV &lt; $p_T^{V}$ &lt; 150 GeV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 jets</td>
</tr>
<tr>
<td>0-lepton</td>
<td>SR</td>
<td>--</td>
</tr>
<tr>
<td>1-lepton</td>
<td>SR plus W + HF CR</td>
<td>--</td>
</tr>
<tr>
<td>2-lepton</td>
<td>SR</td>
<td>$m_{bb}$</td>
</tr>
<tr>
<td>2 lepton</td>
<td>$e\mu$ CR</td>
<td>$m_{bb}$</td>
</tr>
</tbody>
</table>

8.2 Dijet-mass analysis

In the dijet-mass analysis, the $BDT_{VH}$ discriminant is replaced by the $m_{bb}$ variable as the main input used in the global fit, and the number of signal regions is increased from eight to fourteen, as a consequence of splitting the event categories with $p_T^{V}$ > 150 GeV in two in each of the three lepton channels. The different regions entering the likelihood fit are summarised in table 12. Altogether, for the dijet-mass analysis, there are 283 bins in the 18 regions used in the global fit.

8.3 Diboson analysis

The diboson analysis targets diboson production with a $Z$ boson decaying into a pair of $b$-quarks and produced in association with either a $W$ or $Z$ boson. This process has a signature that is similar to the one considered in this analysis, and therefore provides an important validation of the $VH$ result. The cross-section is about nine times larger than for the SM Higgs boson with a mass of 125 GeV, the $m_{bb}$ distribution peaks at lower values, and the $p_T^{bb}$ spectrum is softer. The multivariate discriminant $BDT_{VZ}$ is used to extract the diboson signal. In the diboson-analysis fits, the normalisation of the diboson contributions is allowed to vary with a multiplicative scale factor $\mu_{VZ}$ with respect to the SM prediction, except for the small contribution from $WW$ production, which is treated as a background and constrained within its uncertainty. The overall normalisation uncertainties for the $WZ$ and $ZZ$ processes are removed, while all other systematic uncertainties are kept identical to those in the nominal fit used to extract the Higgs boson signal. 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%. The diboson and Higgs boson $BDT$s provide sufficient separation between the $VZ$ and $VH$ processes that they only have a weak direct correlation (< 1%) in their results.

8.4 Combination with Run 1 data

The statistical analysis of the 13 TeV data is combined with the results of the data recorded at 7 TeV and 8 TeV, reported in ref. [18]. No change is implemented in the analysis of the 7 TeV and 8 TeV data, but several studies were carried out on the correlation and compatibility of the 13 TeV results and the 7 TeV and 8 TeV results. Studies on the correlation of
the experimental systematic uncertainties between the 7 TeV, 8 TeV and 13 TeV analyses were performed for the dominant uncertainties.

The changes in the detector layout (inclusion of the IBL), in the tagging discriminating variable, in the used working points, in the $b$-tagging calibration analyses, and in the way the discriminating variable is used in the analysis support the choice of assuming a negligible correlation in the experimental systematic uncertainties affecting the $b$-tagging across datasets. Nevertheless, even correlating the leading systematic uncertainties for the $b$-jet efficiencies measured in data affects the combined measurement of $\mu$ by less than 5%, and has a negligible impact on its uncertainty. Different correlation schemes for the jet energy scale uncertainties were tested, with no significant impact on the combined result observed.

Studying the impact of potential correlations in the modelling of the background processes is difficult, due to the changes in centre-of-mass energy, Monte Carlo generators, object and event selection, and in the software tools used for simulation, reconstruction and analysis. However, the potential impact of underestimating or omitting correlations is limited by the fact that each of these modelling systematic uncertainties only constitutes a fraction of the total uncertainty, and, furthermore, that this fraction in most cases varies with the centre-of-mass energy following variations in cross-section and acceptance. To evaluate the maximum potential effect of these correlations, a $\chi^2$-combination of the two measurements, the signal strengths from the Run 1 and Run 2 datasets, was performed and studied as a function of different linear correlation coefficients, expressing the degree of correlation between the two measurements. These coefficients were chosen to correspond to different correlation schemes, from uncorrelated to fully correlated, between the $t\bar{t}$, $Z + $HF, and $W + $HF normalisations and systematic shape variations across the two datasets, and they were computed based on the assumed correlation and the relative contribution of a specific uncertainty to the total uncertainty for $\mu$. In all cases considered, the impact on the combined signal strength was found to be smaller than 1%, while the effect on the signal strength uncertainty was found to be smaller than 4%.

As a result of these studies, among the experimental uncertainties, only the $b$-jet-specific jet energy scale uncertainty is correlated across the 7 TeV, 8 TeV and 13 TeV datasets for the combined results. For the Higgs boson signal, theory uncertainties in the overall cross-section, in the $H \rightarrow bb$ branching ratio and in the $p_T^V$-dependent NLO EW corrections, are correlated across the different datasets.

9 Results

The results of the Higgs boson search and diboson analysis are reported below. In the following the fitted signal strength parameters are denoted $\mu$ and $\mu_{VZ}$ rather than $\hat{\mu}$ and $\hat{\mu}_{VZ}$.

9.1 Results of the SM Higgs boson search at $\sqrt{s} = 13$ TeV

Figure 2 shows a selection of characteristic post-fit distributions for each of the lepton channels, while figure 3 shows the BDT output distributions in the most sensitive (high-$p_T^V$) categories. The background prediction in all post-fit distributions is obtained by normalising the backgrounds and setting the systematic uncertainties according to the values of
the floating normalisations and nuisance parameters obtained in the signal extraction fit. The post-global likelihood fit signal and background yields are shown in table 13 for all the analysis regions.

For the tested Higgs boson mass of 125 GeV, when all lepton channels are combined, the probability $p_0$ of obtaining from background alone a result at least as signal-like as the observation is 0.019%. In the presence of a Higgs boson with that mass and the SM signal strength, the expected $p_0$ value is 0.12%. The observation corresponds to an excess with a significance of 3.5 standard deviations, to be compared to an expectation of 3.0 standard deviations. Table 14 shows the $p_0$ and significance values for separate lepton channel fits and for the combined global fit.

For all channels combined the fitted value of the signal strength parameter is

$$\mu = 1.20^{+0.24}_{-0.23} \text{(stat.)}^{+0.34}_{-0.28} \text{(syst.)}. $$

Combined fits are also performed with floating signal strength parameters separately for (i) the three lepton channels, or (ii) the $WH$ and $ZH$ production processes, but leaving all other NPs with the same correlation scheme as for the nominal result. The results of these fits are shown in figures 4 and 5. The compatibility$^5$ of the signal strength parameters measured in the three lepton channels is 10%. The $WH$ and $ZH$ production modes are observed with a significance of 2.4 and 2.6 standard deviations, respectively. The linear correlation term between the signal strengths related to the $WH$ and the $ZH$ production modes is 0.6%. Assuming that the observed signal is due to the SM Higgs boson, with corresponding model-dependent extrapolation corrections to the inclusive phase space, the signal strengths can be interpreted as measurements of the $WH$ and $ZH$ production cross-sections times the $H \rightarrow b\bar{b}$ branching ratio. After removing the theoretical uncertainties for the production cross-sections and branching ratio, these are determined to be

$$\sigma (WH) \times B(H \rightarrow b\bar{b}) = 1.08^{+0.54}_{-0.47} \text{ pb},$$
$$\sigma (ZH) \times B(H \rightarrow b\bar{b}) = 0.57^{+0.26}_{-0.23} \text{ pb},$$

compared to expectations of 0.80 pb and 0.51 pb$^{107}$, respectively. The cross-section for the sum of the $WH$ and $ZH$ production modes is determined to be $\sigma (VH) \times B(H \rightarrow b\bar{b}) = 1.58^{+0.55}_{-0.47} \text{ pb}$, compared to an expectation of 1.31 pb.$^6$ The uncertainties in the quoted theory predictions are negligible compared to the present experimental precision.

Figure 6 shows the data, background and signal yields, where final-discriminant bins in all regions are combined into bins of $\log(S/B)$. Here, $S$ and $B$ are the fitted signal and

$^5$The compatibility of the signal strength across different analysis regions in the fit is evaluated by repeating the fit with different signal strength parameters assigned to each of such $N$ regions, while keeping the rest of the likelihood definition unchanged. Under the hypothesis that the true underlying signal strength parameter values are the same, the difference in the values of profiled $-2 \ln \mathcal{L}$ between the likelihood fit in the nominal and in the new configuration is expected to be distributed according to a $\chi^2$ distribution with number of degrees of freedom equal to $N - 1$. The corresponding $p$-value is thus quoted as a measure of the compatibility.

$^6$The cross-section for the sum of the $WH$ and $ZH$ production modes is obtained from a fit where both production modes are described by a common signal strength parameter. As a result of this, the total cross-section is not equal to the sum of the cross-sections measured for the separate production modes.
Table 13. The fitted Higgs boson signal and background yields for each signal region category in each channel after the full selection of the multivariate analysis. The yields are normalised by the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties. An entry of "\~{}" indicates that a specific background component is negligible in a certain region, or that no simulated events are left after the analysis selection.
Figure 2. The post-fit distributions for $E_{\text{miss}}$ (top left), $m_T^W$ (middle left), $m_T$ (bottom left) and $m_{bb}$ (right) in the 0-lepton (top), 1-lepton (middle) and 2-lepton (bottom) channels for 2-jet, 2-b-tag events in the high $p_T$ region. The background contributions after the global likelihood fit 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 normalised to the signal yield extracted from data ($\mu = 1.20$), and unstacked as an unfilled histogram, scaled by the factor indicated in the legend. The entries in overflow are included in the last bin. The dashed histogram shows the total background as expected from the pre-fit MC simulation. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.
Figure 3. The BDT$_{VH}$ output post-fit distributions in the 0-lepton (top), 1-lepton (middle) and 2-lepton (bottom) channel for 2-b-tag events, in the 2-jet (left) and exactly 3-jet (or ≥ 3 jets for the 2-lepton case) (right) categories in the high $p_T^V$ region. The background contributions after the global likelihood fit 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 normalised to the signal yield extracted from data ($\mu = 1.20$), 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 for the sum of the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel. The BDT$_{VH}$ output distributions are shown with the choice of binning used to define the global likelihood fit.
Figure 4. 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 their combination. The individual $\mu$ values for the lepton channels are obtained from a simultaneous fit with the signal strength parameter for each of the lepton channels floating independently. The compatibility of the individual signal strengths is 10%.

Table 14. The expected and observed $p_0$ and significance values (in standard deviations) for the individual lepton channels and their combination using the 13 TeV dataset. The expected values are evaluated assuming a SM Higgs boson with a mass of 125 GeV.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$p_0$</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-lepton</td>
<td>4.2%</td>
<td>30%</td>
</tr>
<tr>
<td>1-lepton</td>
<td>3.5%</td>
<td>1.1%</td>
</tr>
<tr>
<td>2-lepton</td>
<td>3.1%</td>
<td>0.019%</td>
</tr>
<tr>
<td>Combined</td>
<td>0.12%</td>
<td>0.019%</td>
</tr>
</tbody>
</table>

9.2 Results of the dijet-mass analysis

The distributions of $m_{bb}$ in the dijet-mass analysis are shown in figure 7 for the 2-jet category and the most sensitive analysis regions with $p_T^V > 200$ GeV for the 0-, 1- and 2-lepton channels separately. The $m_{bb}$ distribution for all channels and regions summed, weighted by their respective value of the ratio of fitted Higgs boson signal and background yields, and after subtraction of all backgrounds except for the ($W/Z$)Z diboson processes, is shown in figure 8. The data and the sum of expected signal and backgrounds are found to be in good agreement.
Figure 5. The fitted values of the Higgs boson signal strength parameter $\mu$ for $m_H = 125$ GeV for the $WH$ and $ZH$ processes and their combination. 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. The compatibility of the individual signal strengths is 75%.

Table 15. The numbers of fitted signal and background events and the observed numbers of events in the four highest $S/B$ bins of figure 6. An entry of "-" indicates that a specific background component is negligible in a certain bin, or that no simulated events are left after the analysis selection.
Figure 6. Event yields as a function of log\((S/B)\) for data, background and a Higgs boson signal with \(m_H = 125\) GeV. Final-discriminant bins in all regions are combined into bins of log\((S/B)\), with the fitted signal being \(S\) and the fitted background \(B\). The Higgs boson signal contribution is shown after rescaling the SM cross-section according to the value of the signal strength parameter extracted from data (\(\mu = 1.20\)). The pull (residual divided by its uncertainty) 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 (\(\mu = 1.20\)) and background with respect to the background-only prediction.
Figure 7. The $m_{bb}$ distributions in the 0-lepton (left), 1-lepton (middle) and 2-lepton (right) channels for 2-$b$-tag events, in the 2-jet categories for $p_{T} > 200$ GeV. The background contributions after the global likelihood fit 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 normalised to the signal yield extracted from data ($\mu = 1.30$), 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 for the sum of the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

Figure 8. The distribution of $m_{bb}$ in data after subtraction of all backgrounds except for the $WZ$ and $ZZ$ diboson processes, as obtained with the dijet-mass analysis. The contributions from all lepton channels, $p_{T}$ intervals and number-of-jets categories are summed weighted by their respective value of the ratio of fitted Higgs boson signal and background. The expected contribution of the associated $WH$ and $ZH$ production of a SM Higgs boson with $m_H = 125$ GeV is shown scaled by the measured combined signal strength ($\mu = 1.30$). The size of the combined statistical and systematic uncertainty for the fitted background is indicated by the hatched band.
For all channels combined the fitted value of the signal strength parameter is

\[ \mu = 1.30^{+0.28}_{-0.27} \text{(stat.)}^{+0.27}_{-0.29} \text{(syst.)}, \]

in good agreement with the result of the multivariate analysis. The observed excess has a significance of 3.5 standard deviations, in comparison to an expectation of 2.8 standard deviations. Good agreement is also found in the values of signal strength parameters in the individual channels for the dijet-mass analysis compared to those for the multivariate analysis, with the largest difference between the respective central values of the two analyses being within 15%.

### 9.3 Results of the diboson analysis

The measurement of \(VZ\) production based on the multivariate analysis described in section 8 returns a value of signal strength

\[ \mu_{VZ} = 1.11^{+0.12}_{-0.11} \text{(stat.)}^{+0.22}_{-0.19} \text{(syst.)}, \]

in good agreement with the Standard Model prediction. The \(VZ\) signal is observed with a significance of 5.8 standard deviations, to be compared to an expected significance of 5.3 standard deviations. Analogously to the \(VH\) signal, fits are also performed with separate signal strength parameters for the \(WZ\) and \(ZZ\) production modes, and the results are shown in figure 9. Figure 10 shows the data, background and \(VZ\) signal yields, where final-discriminant bins in all regions are combined into bins of \(\log(S/B)\). Here, \(S\) and \(B\) are the fitted signal and background yields, respectively.

Diboson production is also measured using the dijet-mass analysis. The \(VZ\) signal yield is determined in the fit while the Higgs boson signal yield is kept fixed to the Standard Model prediction within 50% uncertainty. The extracted signal strength is

\[ \mu_{VZ} = 1.01 \pm 0.12 \text{(stat.)}^{+0.20}_{-0.17} \text{(syst.)}, \]

again in good agreement with the Standard Model prediction.

### 9.4 Results of the combination with Run 1

The combination of the Run 1 and Run 2 analyses is used to estimate the combined probability \(p_0\) of obtaining from a background-only experiment a signal at least as large as the one observed, to measure the combined signal strength, and to check the compatibility of the results from the two datasets.

For the tested Higgs boson mass of 125 GeV, the observed \(p_0\) value is 0.018%. In the presence of a Higgs boson with that mass and the SM signal strength, the expected \(p_0\) value is \(3 \times 10^{-5}\). The observation corresponds to an excess with a significance of 3.6 standard deviations, to be compared to an expectation of 4.0 standard deviations. For all channels combined the fitted value of the signal strength parameter is

\[ \mu = 0.90 \pm 0.18 \text{(stat.)}^{+0.21}_{-0.19} \text{(syst.)}. \]
**Figure 9.** The fitted values of the $V_Z$ signal strength parameter $\mu_{VZ}$ for the $WZ$ and $ZZ$ processes and their combination. The individual $\mu_{VZ}$ values for the $(W/Z)Z$ processes are obtained from a simultaneous fit with the signal strength parameters for each of the $WZ$ and $ZZ$ processes floating independently. The compatibility of the individual signal strengths is 88%.
Figure 10. Event yields as a function of $\log(S/B)$ for data, background and $VZ$ processes. Final-discriminant bins in all regions are combined into bins of $\log(S/B)$, with the fitted signal being $S$ and the fitted background $B$. The $VZ$ contribution is shown after rescaling the SM cross-section according to the value of signal strength extracted from data ($\mu = 1.11$). The pull (residual divided by its uncertainty) 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 ($\mu = 1.11$) and background with respect to the background-only prediction.
The Run 1 and Run 2 analyses each contribute three measurements, corresponding to the three lepton channels, yielding a total of six measurements. Their compatibility is estimated to be 7%. Fits are also performed with the signal strength parameters floated independently for the $WH$ and $ZH$ production processes, and for Run 1 and Run 2. The compatibility of the signal strengths for the $WH$ and $ZH$ production processes is 34%, and the results of this fit are shown in figure 11. The compatibility of the signal strength parameters measured in Run 1 with those measured in Run 2 is 21%. Figure 12 shows the signal strengths as measured separately for the 7 TeV, 8 TeV and 13 TeV datasets and their combination. Figure 13 shows the data, background and signal yields, where final-discriminant bins in all regions are combined into bins of $\log(S/B)$. Here, $S$ and $B$ are the fitted signal and background yields, respectively.

10 Conclusion

Evidence for a Standard Model Higgs boson decaying into a $b\bar{b}$ pair and produced in association with a $W$ or $Z$ boson is presented, using data collected by the ATLAS experiment in proton-proton collisions from Run 2 of the Large Hadron Collider. This dataset corresponds to an integrated luminosity of 36.1 fb$^{-1}$ collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. An excess over the expected background is observed, with a significance of 3.5 standard deviations compared to an expectation of 3.0. The measured signal strength with respect to the SM prediction for $m_H = 125$ GeV is found to be $\mu = 1.20^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$.

The analysis procedure adopted to extract the Higgs boson signal is also used to measure the yield of ($W=Z$)$Z$ production with $Z \rightarrow b\bar{b}$, where the ratio of the observed yield to that expected in the Standard Model is found to be $1.11^{+0.12}_{-0.11}(\text{stat.})^{+0.22}_{-0.19}(\text{syst.})$.

The result of the search for the Standard Model Higgs boson based on Run 2 data is combined with previous results based on the full Run 1 dataset collected at centre-of-mass energies of 7 TeV and 8 TeV. An excess over the expected Standard Model background is observed, with a significance of 3.6 standard deviations compared to an expectation of 4.0. The measured signal strength with respect to the SM expectation is found to be $\mu = 0.90 \pm 0.18(\text{stat.})^{+0.21}_{-0.19}(\text{syst.})$. Assuming the Standard Model production strength, the result is consistent with the value of the Yukawa coupling to bottom quarks in the Standard Model.
Figure 11. The fitted values of the Higgs boson signal strength parameter $\mu$ for $m_H = 125$ GeV for the $WH$ and $ZH$ processes and their combination with the 7 TeV, 8 TeV and 13 TeV datasets combined. The individual $\mu$ values for the $(W/Z)H$ processes are obtained from a simultaneous fit with the signal strength parameters for each of the $WH$ and $ZH$ processes floating independently. The compatibility of the individual signal strengths is 34%.

Figure 12. The fitted values of the Higgs boson signal strength parameter $\mu$ for $m_H = 125$ GeV separately for the 7 TeV, 8 TeV and 13 TeV datasets and their combination.
Figure 13. Event yields as a function of log($S/B$) for data, background and Higgs boson signal with $m_H = 125$ GeV, for the 7 TeV, 8 TeV and 13 TeV datasets combined. Final-discriminant bins in all regions are combined into bins of log($S/B$), with the fitted signal being $S$ and the fitted background $B$, for the 7 TeV, 8 TeV and 13 TeV datasets combined. The Higgs boson signal contribution is shown after rescaling the SM cross-section according to the value of signal strength extracted from data ($\mu = 0.90$). The pull (residual divided by its uncertainty) 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 ($\mu = 0.90$) and background with respect to the background-only prediction.
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; BMWFWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; SRNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, 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.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [111].

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


[33] ATLAS collaboration, Summary of ATLAS PYTHIA 8 tunes, ATL-PHYS-PUB-2012-003 [nSPIRE].


INFN e Laboratori Nazionali di Frascati, Frascati, Italy
Fakultät für Mathematik und Physik, Albert-Ludwigs-Universität, Freiburg, Germany
Departement de Physique Nucleaire et Corpusculaire, Université de Genève, Geneva, Switzerland
INFN Sezione di Genova; (b) Dipartimento di Fisica, Università di Genova, Genova, Italy
E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi; (b) High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
II Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany
SUPA - School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
II Physikalisches Institut, Georg-August-Universität, Göttingen, Germany
Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble, France
Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge MA, United States of America
Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany;
Faculty of Applied Information Science, Hiroshima Institute of Technology, Hiroshima, Japan
Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong;
Department of Physics, The University of Hong Kong, Hong Kong; (c) Department of Physics and
Institute for Advanced Study, The Hong Kong University of Science and Technology, Clear Water
Bay, Kowloon, Hong Kong, China
Department of Physics, National Tsing Hua University, Hsinchu, Taiwan
Department of Physics, Indiana University, Bloomington IN, United States of America
Institut für Astro- und Teilchenphysik, Leopold-Franzens-Universität, Innsbruck, Austria
University of Iowa, Iowa City IA, United States of America
Department of Physics and Astronomy, Iowa State University, Ames IA, United States of America
Joint Institute for Nuclear Research, JINR Dubna, Dubna, Russia
KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
Graduate School of Science, Kobe University, Kobe, Japan
Faculty of Science, Kyoto University, Kyoto, Japan
Kyoto University of Education, Kyoto, Japan
Research Center for Advanced Particle Physics and Department of Physics, Kyushu University,
Pakuoka, Japan
Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata, Argentina
Physics Department, Lancaster University, Lancaster, United Kingdom
INFN Sezione di Lecce; (b) Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy
Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
Department of Experimental Particle Physics, Jožef Stefan Institute and Department of Physics,
University of Ljubljana, Ljubljana, Slovenia
School of Physics and Astronomy, Queen Mary University of London, London, United Kingdom
Department of Physics, Royal Holloway University of London, Surrey, United Kingdom
Department of Physics and Astronomy, University College London, London, United Kingdom
Louisiana Tech University, Ruston LA, United States of America
Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and
CNRS/IN2P3, Paris, France
Fysiska institutionen, Lunds universitet, Lund, Sweden
Departamento de Física Teórica C-15, Universidad Autonoma de Madrid, Madrid, Spain
Institut für Physik, Universität Mainz, Mainz, Germany
School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France
Department of Physics, University of Massachusetts, Amherst MA, United States of America
Department of Physics, McGill University, Montreal QC, Canada
Institute of Physics, Academy of Sciences of the Czech Republic, Praha, Czech Republic
Czech Technical University in Prague, Prague, Czech Republic
Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
State Research Center Institute for High Energy Physics (Protvino), NRC KI, Russia
Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
\(^{(a)}\) INFN Sezione di Roma; \(^{(b)}\) Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy
\(^{(c)}\) INFN Sezione di Roma Tor Vergata; \(^{(d)}\) Dipartimento di Fisica, Università di Roma Tor Vergata, Roma, Italy
\(^{(e)}\) INFN Sezione di Roma Tre; \(^{(f)}\) Dipartimento di Matematica e Fisica, Università Roma Tre, Roma, Italy

\(^{(a)}\) Faculté des Sciences Ain Chock, Réseau Universitaire de Physique des Hautes Energies - Université Hassan II, Casablanca; \(^{(b)}\) Centre National de l’Energie des Sciences Techniques Nucléaires, Rabat; \(^{(c)}\) Faculté des Sciences Semlalia, Université Cadi Ayyad, LPHEA-Marrakech; \(^{(d)}\) Faculté des Sciences, Université Mohamed Premier and LPTPM, Oujda; \(^{(e)}\) Faculté des sciences, Université Mohammed V, Rabat, Morocco

DSM/IRFU (Institut de Recherches sur les Lois Fondamentales de l’Univers), CEA Saclay (Commissariat à l’Energie Atomique et aux Energies Alternatives), Gif-sur-Yvette, France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz CA, United States of America
Department of Physics, University of Washington, Seattle WA, United States of America
Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
Department of Physics, Shinshu University, Nagano, Japan
Department Physik, Universität Siegen, Siegen, Germany
Department of Physics, Simon Fraser University, Burnaby BC, Canada
SLAC National Accelerator Laboratory, Stanford CA, United States of America

\(^{(a)}\) Department of Physics, University of Cape Town, Cape Town; \(^{(b)}\) Department of Physics, University of Johannesburg, Johannesburg; \(^{(c)}\) School of Physics, University of the Witwatersrand, Johannesburg, South Africa

\(^{(a)}\) Department of Physics, Stockholm University; \(^{(b)}\) The Oskar Klein Centre, Stockholm, Sweden
Physics Department, Royal Institute of Technology, Stockholm, Sweden
Departments of Physics & Astronomy and Chemistry, Stony Brook University, Stony Brook NY, United States of America
Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom
School of Physics, University of Sydney, Sydney, Australia
Institute of Physics, Academia Sinica, Taipei, Taiwan
Department of Physics, Technion: Israel Institute of Technology, Haifa, Israel
Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
International Center for Elementary Particle Physics and Department of Physics, The University of Tokyo, Tokyo, Japan
Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
Tomsk State University, Tomsk, Russia
Department of Physics, University of Toronto, Toronto ON, Canada
\(^{(a)}\) INFN-TIFPA; \(^{(b)}\) University of Trento, Trento, Italy
\(^{(c)}\) TRIUMF, Vancouver BC; \(^{(d)}\) Department of Physics and Astronomy, York University, Toronto ON, Canada
Faculty of Pure and Applied Sciences, and Center for Integrated Research in Fundamental Science and Engineering, University of Tsukuba, Tsukuba, Japan
Department of Physics and Astronomy, Tufts University, Medford MA, United States of America
Department of Physics and Astronomy, University of California Irvine, Irvine CA, United States of America
INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine; ICTP, Trieste; Dipartimento di Chimica, Fisica e Ambiente, Università di Udine, Udine, Italy
Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
Department of Physics, University of Illinois, Urbana IL, United States of America
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia - CSIC, Spain
Department of Physics, University of British Columbia, Vancouver BC, Canada
Department of Physics and Astronomy, University of Victoria, Victoria BC, Canada
Department of Physics, University of Warwick, Coventry, United Kingdom
Waseda University, Tokyo, Japan
Department of Particle Physics, The Weizmann Institute of Science, Rehovot, Israel
Department of Physics, University of Wisconsin, Madison WI, United States of America
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität, Würzburg, Germany
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
Department of Physics, Yale University, New Haven CT, United States of America
Yerevan Physics Institute, Yerevan, Armenia
Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Villeurbanne, France
Academia Sinica Grid Computing, Institute of Physics, Academia Sinica, Taipei, Taiwan

\( ^{a} \) Also at Department of Physics, King’s College London, London, United Kingdom
\( ^{b} \) Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
\( ^{c} \) Also at Novosibirsk State University, Novosibirsk, Russia
\( ^{d} \) Also at TRIUMF, Vancouver BC, Canada
\( ^{e} \) Also at Department of Physics & Astronomy, University of Louisville, Louisville, KY, United States of America
\( ^{f} \) Also at Physics Department, An-Najah National University, Nablus, Palestine
\( ^{g} \) Also at Department of Physics, California State University, Fresno CA, United States of America
\( ^{h} \) Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
\( ^{i} \) Also at II Physikalisches Institut, Georg-August-Universität, Göttingen, Germany
\( ^{j} \) Also at Departamento de Fisica de la Universitat Autonoma de Barcelona, Barcelona, Spain
\( ^{k} \) Also at Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Portugal
\( ^{l} \) Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
\( ^{m} \) Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing, China
\( ^{n} \) Also at Università di Napoli Parthenope, Napoli, Italy
\( ^{o} \) Also at Institute of Particle Physics (IPP), Canada
\( ^{p} \) Also at Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania
\( ^{q} \) Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
\( ^{r} \) Also at Borough of Manhattan Community College, City University of New York, New York City, United States of America
\( ^{s} \) Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece
\( ^{t} \) Also at Centre for High Performance Computing, CSIR Campus, Rosebank, Cape Town, South Africa
\( ^{u} \) Also at Louisiana Tech University, Ruston LA, United States of America
\( ^{v} \) Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain