A search for the dimuon decay of the Standard Model Higgs boson in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector

The ATLAS Collaboration

A search for the dimuon decay of the Standard Model Higgs boson is performed using data corresponding to an integrated luminosity of 139 fb$^{-1}$ collected in Run 2 with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 13$ TeV at the Large Hadron Collider. No significant excess is observed above the expected background. The observed upper limit on the cross section times branching ratio is $1.7\times$ the Standard Model prediction at 95% confidence level for a Higgs boson mass of 125.09 GeV, while the expected limit in the absence (assuming SM signal yield) of a $H \rightarrow \mu\mu$ signal is $1.3\times(2.2)$. The best fit value of the signal strength parameter, defined as the ratio of the observed signal yield to the one expected in the Standard Model, is $\mu = 0.5 \pm 0.7$.

Update July 23 2019: a mismatch between category names and spurious signal values was corrected in Table 4. The change only impacts this Table while all results are unchanged.
1 Introduction

In July 2012, the ATLAS and CMS Collaborations announced the discovery of a new particle with a mass of approximately 125 GeV [1, 2] at the CERN Large Hadron Collider (LHC). Subsequent measurements have indicated that this particle is consistent with the Standard Model (SM) Higgs boson [3–6], denoted by \( H \).

While the interaction between the Higgs boson and the charged fermions of the third-generation has already been observed by both the ATLAS and CMS Collaborations [6–10], only upper limits exist on the interactions with fermions of the other generations. The \( H \rightarrow \mu\mu \) decay offers the best opportunity to measure the Higgs-boson interactions with the second-generation fermions at the LHC. The SM branching ratio to dimuons for the Higgs boson with \( m_H = 125.09 \text{ GeV} \) is \( 2.17 \times 10^{-4} \) [11]. However, physics beyond the SM [12–14] could modify the branching ratio. Any deviation from the SM prediction could be a sign of new physics.

Both the ATLAS and CMS Collaborations carried out searches for the \( H \rightarrow \mu\mu \) process based on partial sets of the data collected during the LHC Run 2 [15, 16]. In this note, an update of the search for the dimuon decay of the Higgs boson is presented using \( pp \) collision data recorded with the ATLAS detector in the full LHC Run 2 period, spanning from 2015 to 2018 at \( \sqrt{s} = 13 \text{ TeV} \), corresponding to an integrated luminosity of about 139 fb\(^{-1}\). The Higgs boson is assumed to have a mass of \( m_H = 125.09 \text{ GeV} \) [5, 17] for all results presented.

The analysis selects events with two opposite-charge muons and classifies them into twelve mutually exclusive categories on the basis of multivariate discriminants. The discriminants are defined to exploit the kinematic differences between the main Higgs boson production modes, gluon fusion (\( ggF \)) and vector boson fusion (VBF), and the background processes dominated by the Drell–Yan (DY) process \( \gamma^* \rightarrow \mu\mu \). Signal events are characterised by a more central dimuon system with harder transverse momentum (\( p_T \)). The jet multiplicity is expected to be higher in both \( ggF \) and VBF production with respect to the DY background. In addition, the VBF process has a unique signature with two high \( p_T \) jets with a large rapidity gap and with little hadronic activity between them.

With respect to the analysis described in Ref. [15], several improvements are incorporated: the classification of the events in different jet multiplicities with dedicated multivariate discriminants, a better modeling of the background based on a large increase of the DY simulation dataset, the use of reconstructed photons to improve the signal mass resolution for events with final state radiation and increased rejection of jets from pile-up in the forward region.

After event categorisation, the signal yield is extracted by a simultaneous fit to the twelve dimuon mass (\( m_{\mu\mu} \)) distributions in the range 110–160 GeV together with background normalisation and shape parameters, exploiting the resonant behavior of the Higgs boson signal.
2 ATLAS Detector

The ATLAS detector [18, 19] covers nearly the entire solid angle around the collision point.\textsuperscript{1} 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 toroid magnets.

The inner detector (ID) system is immersed in a 2 T axial magnetic field and provides charged particle tracking in the range $|\eta| < 2.5$. A high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track. It is followed by a silicon microstrip tracker, which usually provides four two-dimensional measurement points per track. These silicon detectors are complemented by a transition radiation tracker, which enables radially extended track reconstruction up to $|\eta| < 2.0$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity liquid-argon (LAr) electromagnetic 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 scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two LAr hadronic endcap calorimeters.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by superconducting air-core toroids. The precision chamber system covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

3 Data and MC samples

The $pp$ collision data at $\sqrt{s} = 13$ TeV analyzed here correspond to the full Run 2 dataset, with an integrated luminosity of 139 fb$^{-1}$ after the application of data quality requirements. The mean number of $pp$ interactions per bunch crossing was about 34. Events used in this analysis were recorded using a combination of single-muon triggers with transverse momentum thresholds up to 26 GeV for isolated muons and 50 GeV for muons without any isolation requirement imposed. The trigger efficiency is about 91% for the signal processes with respect to the event selection discussed in Section 4.

Samples of simulated Monte Carlo (MC) events are used to optimise the selection, to model the signal processes and to develop an analytic function to model the $m_{\mu\mu}$ distributions for the background estimate. The signal samples as well as a complete set of background processes as discussed in the following were processed through the full ATLAS detector simulation [20] based on Geant4 [21].

Signal samples were generated for the main Higgs boson production modes. The mass of the Higgs boson was set in the simulation to $m_H = 125$ GeV and the corresponding width is $\Gamma_H = 4.07$ MeV [22]. The samples are normalised with the latest available theoretical calculations of the corresponding SM production

\textsuperscript{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 along the beam pipe. The x-axis points from the IP to 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 rapidity is defined as $y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$ and the distance between two objects is defined as $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$.

3
cross sections, summarised in Ref. [11]. The normalization of all Higgs boson samples also accounts for the $H \rightarrow \mu \mu$ branching ratio of $2.18 \times 10^{-4}$ calculated with HDECAY [23–25] and PROPHECY4F [26–28].

Higgs boson production in the $ggF$ process is simulated using the Powheg NNLOPS program [29–36] with the PDF4LHC15 set of parton distribution functions (PDFs) [37]. The simulation achieves next-to-next-to-leading-order (NNLO) accuracy in QCD for inclusive observables after reweighting the Higgs boson rapidity spectrum [38]. The parton-level events are processed by Pythia8 [39] to decay the Higgs bosons and to provide parton showering, final state photon radiation (QED FSR), hadronisation and underlying event, using the AZNLO set of tuned parameters [40]. The sample is normalised to a next-to-next-to-next-to-leading-order (NNLO) electroweak corrections [41–52].

Higgs boson production via VBF and $q\bar{q}/qg \rightarrow VH$ are generated at NLO accuracy in QCD using the Powheg-Box program [53–55]. The loop-induced process $gg \rightarrow ZH$ is generated at leading order using Powheg-Box. For the VBF and $VH$ samples the same settings for the PDF set and Pythia8 as in the $ggF$ sample are employed. The VBF sample is normalised to an approximate-NNLO QCD cross section with NLO electroweak corrections [56–58]. The $VH$ samples are normalised to cross sections calculated at NNLO in QCD with NLO electroweak corrections for $q\bar{q} \rightarrow VH$ and at NLO and next-to-leading-logarithm accuracy in QCD for $gg \rightarrow ZH$ [59–66]. Higgs boson production via $t\bar{t}H$ is generated at NLO accuracy in QCD using MG5_aMC@NLO [67, 68] with the NNPDF3.0NLO PDF set [69] and interfaced to Pythia8 using the A14 parameter set [70]. The cross section is taken from a calculation accurate to NLO in QCD with NLO electroweak corrections [71–74].

DY background events are generated with the Sherpa v2.2 [75] generator using NLO-accurate matrix elements for emissions of up to two jets, and LO-accurate matrix elements for emissions of up to four jets calculated with the Comix [76] and OpenLoops [77, 78] libraries and the NNPDF3.0 NNLO set. They are matched with the Sherpa parton shower [79] using the MEPS@NLO prescription [80–83]. Diboson processes ($WW$, $WZ$, and $ZZ$) as well as electroweak $Zjj$ production are generated in a similar setup with Sherpa v2.2. The $t\bar{t}$ and single-top quark samples are generated with Powheg-Box [84, 85] using the NNPDF3.0NLO PDF set interfaced to Pythia8 for parton showering and hadronisation using the A14 parameter set.

The effects from multiple $pp$ collisions in the same or neighboring bunch crossings (pileup) are included in the MC simulation by overlaying inelastic $pp$ interactions produced using Pythia8 using the NNPDF2.3LO set of PDFs [86] and the A3 tune [87]. Events are reweighted such that the distribution of the average number of interactions per bunch crossing matches that observed in data. Simulated events are corrected to reflect the muon and jet momentum scales and resolutions as well as the muon trigger, reconstruction, identification, and isolation efficiencies measured in data.

As the background samples discussed above do not provide the statistical accuracy to test and refine the background modelling at the level required to detect the $H \rightarrow \mu \mu$ signal, a dedicated fast simulation was developed for the dominant DY background. For this sample $Z/\gamma^* + 0, 1$ jet events are generated inclusively at NLO accuracy using Powheg [88] with the CT10 PDF set [89]. Additional $Z/\gamma^* + 2$ jet events are generated with Alpgen [90] at LO accuracy with the CTEQ6L1 PDF set [91]. The events are interfaced to Photos [92] to simulate QED FSR. All further processing steps are approximated using fast parameterisations extracted from fully simulated MC samples or directly from ATLAS data. These parameterisations approximate the effects of parton showering or experimental detection efficiencies by weighting of events or using pre-determined probability distributions to model the response of the ATLAS detector for the objects used in the analysis as discussed in Section 4. The parameterisation of the ATLAS
detector comprises a detailed description of the muon momentum resolution and muon efficiencies, photons from QED FSR, hadronic jets from the primary interaction and pileup events, and effect of pileup and underlying event on the measurement of the missing transverse momentum $E_T^{\text{miss}}$. In total about 20 billion events are prepared this way, corresponding to an equivalent luminosity of about 100 ab$^{-1}$ in the region relevant for the analysis.

4 Object definitions and event selection

Events are required to contain at least one reconstructed $pp$ collision vertex candidate with at least two associated ID tracks each with $p_T > 0.5$ GeV. The vertex with the largest sum of $p_T^2$ of tracks is considered to be the primary vertex of the hard interaction.

The majority of muon candidates is reconstructed by combining tracks in the ID with tracks in the MS. To improve the muon reconstruction efficiency in the region of $|\eta| < 0.1$, which has limited coverage in the MS, additional muon candidates are identified by matching a fully reconstructed ID track to either an MS track segment or a calorimetric energy deposit consistent with a minimum-ionising particle. In the region $2.5 < |\eta| < 2.7$, which is not covered by the ID, additional muons are reconstructed from MS tracks with hits in the three MS layers and combined with forward ID hits, if possible. Muon candidates are required to satisfy the “loose” criteria defined in Ref. [93] and have $p_T > 15$ GeV and $|\eta| < 2.7$. Muons with an associated ID track are matched to the primary vertex with a longitudinal impact parameter ($|d_0|/\sigma(d_0)<3$, where $\sigma(d_0)$ is the uncertainty in $d_0$). Furthermore, isolation criteria are applied using information on ID tracks and calorimeter deposits in a range $\Delta R < 0.2$ around the muon to suppress non-prompt muons originating from hadron decays.

As muons may lose a significant fraction of their energy by QED FSR, up to one final-state photon candidate is included in the $m_{\mu\mu}$ calculation to improve the signal reconstruction. Photon candidates are reconstructed with a procedure similar to the one described in Ref. [94]. The procedure is optimised to achieve the best sensitivity for the $H \rightarrow \mu\mu$ signal. For photon candidates close to muons $\Delta R(\gamma,\mu) < 0.2$, a variable threshold is applied that increases linearly from $p_T^\gamma > 3$ GeV for $\Delta R = 0$ to $p_T^\gamma > 8$ GeV for $\Delta R = 0.2$. Photons away from muons by $\Delta R(\gamma,\mu) > 0.2$ have to pass tight identification requirements and $p_T^\gamma > 10$ GeV [95]. A QED FSR candidate is found in about 5% of the events and the signal $m_{\mu\mu}$ width is reduced by about 3% when considering all reconstructed signal events.

Jets are reconstructed from topologically clustered energy deposits in the calorimeters [96] using the anti-$k_T$ algorithm [97, 98] with a radius parameter of $R = 0.4$. Candidate jets must have $|\eta| < 4.5$, and the jet $p_T$ must be larger than 25 (30) GeV for $|\eta| < 2.5$ ($2.5 < |\eta| < 4.5$). To suppress pileup contributions, jets with $|\eta| < 2.5$ and $p_T < 120$ GeV that do not originate from the primary vertex are rejected using the jet vertex tagging algorithm (JVT) [99], which combines tracking information into a multivariate likelihood. A suppression of pile-up jets in the forward region $2.5 < |\eta| < 4.5$ is achieved with the fJVT discriminant [100], which exploits the jet shape and the topological correlations in pile-up interactions.

Jets containing $b$ hadrons with $|\eta| < 2.5$ are identified as $b$-tagged jets using a multivariate $b$-tagging algorithm that provides a 60% efficiency and a rejection factor of more than 1000 for light-flavor jets [101, 102].
Table 1: Summary of the main event selection criteria common to all events (top) as well as the criteria applied to the selection of hadronic jets (bottom). The middle section defines three ranges in the dimuon invariant mass $m_{\mu\mu}$ used in the analysis, as described in the text.

<table>
<thead>
<tr>
<th></th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>Primary vertex</td>
</tr>
<tr>
<td></td>
<td>Two opposite-charge muons</td>
</tr>
<tr>
<td></td>
<td>$</td>
</tr>
<tr>
<td></td>
<td>No $b$-tagged jets</td>
</tr>
<tr>
<td>$Z$ region</td>
<td>$76 &lt; m_{\mu\mu} &lt; 106$ GeV</td>
</tr>
<tr>
<td>Sideband region</td>
<td>$110 &lt; m_{\mu\mu} &lt; 120$ GeV or $130 &lt; m_{\mu\mu} &lt; 180$ GeV</td>
</tr>
<tr>
<td></td>
<td>$110 &lt; m_{\mu\mu} &lt; 160$ GeV</td>
</tr>
<tr>
<td>Fit region</td>
<td>$p_T &gt; 25$ GeV and $</td>
</tr>
<tr>
<td></td>
<td>or with $p_T &gt; 30$ GeV and $2.5 &lt;</td>
</tr>
<tr>
<td>Jets</td>
<td></td>
</tr>
</tbody>
</table>

Neutrinos escape from the detector and lead to missing transverse momentum $E_T^{\text{miss}}$. The $E_T^{\text{miss}}$ is defined as the magnitude of the negative vectorial sum of the transverse momenta of the selected and calibrated physics objects (including muons and jets) and the ID tracks not associated with any physics object (soft term) [103].

Events are selected if they contain two opposite-charge muon candidates. The leading muon is required to have $p_T > 27$ GeV to be above the trigger threshold. Top-quark production is the second largest background and is suppressed by vetoing events containing $b$-tagged jets. This selection provides a total acceptance times efficiency of about 59% for the ggF and VBF signal events. The main selection requirements are summarised in Table 1, which also defines three ranges in the dimuon invariant mass $m_{\mu\mu}$ considered in the analysis for different purposes. Almost 100 million events are selected in data in the $Z$-region and are used to test the detector performance. About 2 million data events in the signal-depleted sideband region are used to train multivariate classifiers and to test the background modelling. The region $m_{\mu\mu} = 120$–130 GeV contains most of the $H \rightarrow \mu\mu$ signal with about 860 signal events expected as part of the 450,000 selected data events. The final signal+background fits are performed in the region $m_{\mu\mu} = 110$–160 GeV, which allows to constrain the backgrounds at the same time as extracting the signal.

5 Event categorisation

Events passing the preselection criteria of Section 4 are first classified depending on the number of reconstructed jets into three categories: 0-jet, 1-jet and 2-jet, where the latter includes events with two or more jets.

To fully exploit the kinematic difference between the signal and the backgrounds, which are dominated by DY dimuon production contributing more than 90% to the total background after preselection, boosted decision trees (BDT) [104, 105] are trained in each jet multiplicity category with the XGBoost package [106]. All BDTs are trained using the data events in the sideband regions as background and $H \rightarrow \mu\mu$ simulation as signal as specified below. For the background processes, the distributions of the variables used in the BDT training are very similar between the mass sideband and signal regions as shown in Figure 1. In order to avoid any potential bias, all trainings are performed using a four-fold method, where for each fold
50% of events are used for training, 25% for validation and 25% for testing. The validation fold is used to optimise the BDT hyper-parameters while the test is used to evaluate the BDT performance. In order to match BDT scores across the four folds, they are transformed so that the BDT scores of the combined ggF and VBF samples has a uniform distribution in each of the folds. After the transformation of the BDT outputs, the test sets of all the four folds are then combined together for the final evaluation.

In the 2-jet category, a first BDT is trained to disentangle signal events produced by VBF from background events. This BDT, with a score referred to as \(O_{\text{VBF}}\), is based on the following 14 variables related to the dijet and dimuon systems. The dijet system is characterised by the transverse momentum \(p_T^{j}\), rapidity \(y_{jj}\) and the absolute value of \(\cos \theta^{*}\) in the Collins–Soper frame [107]. In both ggF and VBF production the signal events are characterised by larger \(p_T^{\mu\mu}\) and smaller absolute values of \(y_{\mu\mu}\) with respect to the dominant DY background, as well as smaller values of \(|\cos \theta^{*}|\). For the leading and subleading jets in the event (denoted as \(j_1\) and \(j_2\)), the following variables are computed: \(p_T\) and \(\eta\) of \(j_1\) and \(j_2\), the azimuthal-angle differences between each jet and the dijet system \(\Delta \phi_{\mu\mu,j_1}\) and \(\Delta \phi_{\mu\mu,j_2}\), the kinematics of the dijet system \((jj)\) characterised by \(p_T^{jj}\), mass \(m_{jj}\) and \(y_{jj}\), and the azimuthal difference between the dijet and the dimuon systems \(\Delta \phi_{\mu\mu,jj}\). In addition, the missing transverse momentum \(E_T^{\text{miss}}\), which can discriminate the \(t\bar{t}\) background from the signal, is also considered. The expected distributions of these 14 variables for the simulated \(H \rightarrow \mu\mu\) signal and background and for events observed in the data sidebands are depicted in Figure 1. From the \(O_{\text{VBF}}\) classifier three categories in the region with the highest score are selected.

The remaining events are split by jet multiplicity and considered for further classification in three BDTs, with scores denoted as \(O_{\text{ggF}}^{(0-2)}\). These BDTs are trained with the \(H \rightarrow \mu\mu\) ggF and VBF production as signal. In each jet category, the same three variables characterising the dijet and dimuon systems discussed above are used, i.e. \(p_T^{\mu\mu}\), \(y_{\mu\mu}\) and \(|\cos \theta^{*}|\). They are complemented by the transverse momentum \(p_T^{\mu j}\), pseudo-rapidity \(\eta_{jj}\) and the azimuthal-angle difference to the dimuon system \(\Delta \phi_{\mu\mu,j1}\) for the 1-jet and \(\Delta \phi_{\mu\mu,j2}\) for the 0-jet BDT. In the 2-jet category the same variables as used for the VBF BDT are used. The events in each of the three jet multiplicity categories are further classified into three categories on the basis of the \(O_{\text{ggF}}^{(0-2)}\) scores.

In total, the events are thus classified into twelve mutually exclusive categories summarised in Table 2: three for 2-jet events on the basis of the \(O_{\text{VBF}}\) score and nine on the basis of the \(O_{\text{ggF}}^{(0-2)}\) scores (three for each jet multiplicity). The BDT score boundaries are chosen by maximising the sensitivity to the SM \(H \rightarrow \mu\mu\) signal. In each of the twelve BDT categories the signal is extracted with a fit to the data mass spectrum in the invariant mass range \(m_{\mu\mu} = 110–160\) GeV, as described in Section 6.

<table>
<thead>
<tr>
<th>Category</th>
<th>0-jet</th>
<th>1-jet</th>
<th>VBF</th>
<th>2-jet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(O_{\text{ggF}}^{0}) ≥ 0.75</td>
<td>(O_{\text{ggF}}^{1}) ≥ 0.78</td>
<td>(O_{\text{VBF}}) ≥ 0.89</td>
<td>(O_{\text{ggF}}^{2}) ≥ 0.48</td>
</tr>
<tr>
<td>High</td>
<td>(O_{\text{ggF}}^{0}) &lt; 0.75</td>
<td>0.38 ≤ (O_{\text{ggF}}^{1}) &lt; 0.78</td>
<td>0.77 ≤ (O_{\text{VBF}}) &lt; 0.89</td>
<td>0.22 ≤ (O_{\text{ggF}}^{2}) &lt; 0.48</td>
</tr>
<tr>
<td>Medium</td>
<td>(O_{\text{ggF}}^{0}) &lt; 0.35</td>
<td>(O_{\text{ggF}}^{1}) &lt; 0.38</td>
<td>0.60 ≤ (O_{\text{VBF}}) &lt; 0.77</td>
<td>(O_{\text{ggF}}^{2}) &lt; 0.22</td>
</tr>
<tr>
<td>Low</td>
<td>(O_{\text{ggF}}^{0}) &lt; 0.35</td>
<td>(O_{\text{ggF}}^{1}) &lt; 0.38</td>
<td>0.60 ≤ (O_{\text{VBF}}) &lt; 0.77</td>
<td>(O_{\text{ggF}}^{2}) &lt; 0.22</td>
</tr>
</tbody>
</table>
Figure 1: The distributions of the 14 variables used in the VBF BDT for $n_j \geq 2$ events for $H \rightarrow \mu\mu$ signal MC (red), data sideband (black) and background MC sideband and center (defined as 120-130 GeV mass range). The background MC includes DY from full simulation, $t\bar{t}$, diboson, single top-quark events.
6 Signal and Background Modeling and Systematic Uncertainties

The signal extraction is based on a binned maximum-likelihood fit to the invariant mass spectrum of the dimuon system as described in Section 7. Analytic models are used in the fit to describe the $m_{\mu\mu}$ distributions for both the signal and background processes.

6.1 Signal modelling

In the SM, the $H \rightarrow \mu\mu$ signal is predicted to be a narrow resonance with a width of 4.1 MeV for $m_H = 125.09$ GeV. The observed signal shape is thus determined by detector resolution effects on the muon momentum measurement. The functional form used for the Higgs signal model is a double-sided Crystal Ball. This function is a modification of the Crystal-Ball function [108, 109], consisting of a Gaussian central part and power-law tails on both sides.

For each BDT category, the signal parameters are fitted both individually for each signal production mode ($ggF$, VBF, $VH$, $t\bar{t}H$) and summing all production modes assuming relative normalisations between them as predicted by the SM. No significant differences were found and the signal shape is hence described using inclusive parameterisations, derived separately for each of the twelve BDT categories by a fit to the signal MC mass spectrum. The value of the width of the Gaussian component of the double-sided Crystal Ball function varies between 2.5 and 3.0 GeV depending on the category. Potential biases on the extracted signal yields due to the analytic parameterisations are tested with a signal injection procedure. A signal plus background fit is performed in each category on pseudo-data obtained adding the MC signal to the expected background distribution obtained from the background functions described in the following. In each category, the observed values of the extracted signal yields are found to be in agreement with those injected within the statistical accuracy of the test of about 0.5%.

Several sources of systematic uncertainties on the signal modelling are considered, including both theoretical and experimental effects. The theoretical uncertainties on the signal production affect the number of signal events expected in each BDT category. The uncertainties considered for the main production modes ($ggF$ and VBF) include the impact of the missing higher order QCD corrections, PDFs and underlying event and hadronisation. In particular, the uncertainty on the $ggF$ signal is derived based on the approach described in Ref. [11], including effects from the variation of QCD scales for factorization, renormalisation and resummation, and the migration between jet-multiplicity regions. The uncertainty on the $ggF$ Higgs boson transverse momentum, including migration effect between different kinematic regions as well as the effect of the treatment of the top-quark mass in the loop corrections, is also taken into account. In addition, dedicated uncertainties are assigned for $ggF$ signal acceptance in VBF topologies. The uncertainty on the predicted SM branching ratio and Higgs boson production cross sections are included according to Ref. [11]. The uncertainties associated with the modeling of the underlying event and parton showering are estimated by considering the Pythia8 systematic eigentune variations and by comparing events showered by Pythia8 with those showered by Herwig7 [110, 111]. The impact of the theory uncertainties on the predicted signal yield acceptances in the different categories ranges between a few per mill and 15% for the $ggF$ production. Similarly for VBF production the impact of the theory uncertainties on the predicted acceptances varies between a few per mill and 7%.

Systematic uncertainties related to the different reconstructed physics objects used in the analysis affect the expected signal yields in each BDT category. In addition, systematic uncertainties in the muon momentum scale and resolution affect also the signal mass distribution. The experimental uncertainties considered
are the muon reconstruction and identification efficiencies, the efficiencies due to the trigger, isolation and impact parameter requirements, the muon momentum scale and resolution, the determination of the $E^\text{miss}_T$ soft term, the $b$-tagging efficiency, the pileup modeling, as well as the jet energy scale and resolution. The impact of the experimental uncertainties on the predicted signal yields and modeling in the different categories are dominated by the uncertainties in the jet energy scale and resolution and muon momentum resolution. The former can affect signal yields up to about 10% in some of the 2-jet categories. The muon momentum resolution uncertainty has an impact on the fitted yields ranging between 1 and 6% depending on the category.

The experimental uncertainty in the assumed value of the Higgs mass $m_H = 125.09 \text{ GeV}$ from [5, 17] is also taken into account. All these sources of uncertainties are included in the signal extraction fit described in Section 7 in terms of nuisance parameters acting on the relative signal yields in the different BDT categories and on the signal mass distributions.

### 6.2 Background modelling

Due to the very small signal to background ratio, about 0.2% in the region $m_{\mu\mu} = 120–130 \text{ GeV}$ for an inclusive selection, the background determination is of paramount importance. The $m_{\mu\mu}$ background spectrum is parameterised by analytical functions that can describe this distribution at the per mill level to avoid a significant bias on the extracted signal yields.

The background is dominated by the DY process, which accounts for more than 90% of the total, with small contributions from top-quark processes (mainly in the 2-jet categories) and diboson production.

To achieve the required accuracy in the analytic description of the background $m_{\mu\mu}$ distribution, the following approach is used. A core function that describes the DY mass shape inclusively is multiplied by an empirical function that can correct for distortions of the mass shape due the event selection and categorization and other smaller background contributions. The core function has no free parameters and is common to all categories while the empirical functions have a certain number of free parameters that are fit to data independently in each category.

The core component of the background function is based on the LO DY line-shape (see e.g. Ref. [112]):

$$DY(m_{\mu\mu}) = \sum_q L_{q\bar{q}}(m_{\mu\mu}) \cdot \sigma_{q\bar{q}}(m_{\mu\mu}), \ q = u, \ s, \ d.$$  \hspace{1cm} (1)

The parton luminosity contribution $L_{q\bar{q}}$ in Eq. 1 is derived from PDF4LHC15 as a function of $\hat{s} = m_{\mu\mu}^2$ for the LO DY case using APFEL [113] interfaced to LHAPDF [114] and parameterised using a 6th order polynomial. The matrix element component $\sigma_{q\bar{q}}(\hat{s}) = \sigma_{q\bar{q}}(m_{\mu\mu})/(2m_{\mu\mu})$ can be expressed as

$$\sigma_{q\bar{q}}(\hat{s}) = \frac{4\pi\alpha^2}{3\hat{s}N_c}[Q_q^2 - 2Q_qV_qV_{\bar{q}}\chi_{Z\gamma}(\hat{s}) + (A_{1q}^2 + V_{1q}^2)(A_{1\bar{q}}^2 + V_{1\bar{q}}^2)\chi_{Z}(\hat{s})].$$
where

$$
\chi_Z(\hat{s}) = \kappa \frac{\hat{s}(\hat{s} - m_Z^2)}{(\hat{s} - m_Z^2)^2 + \Gamma_Z^2 m_Z^2},
$$

$$
\chi_{\gamma}(\hat{s}) = \kappa^2 \frac{(\hat{s} - m_Z^2)^2}{(\hat{s} - m_Z^2)^2 + \Gamma_Z^2 m_Z^2},
$$

$$
\kappa = \sqrt{\frac{2G_F m_Z^2}{4\pi\alpha}}.
$$

Here $Q, V, A$ denote the electric charges, vector and axial-vector couplings of the fermions, $\alpha, G_F$ the electroweak couplings, $m_Z, \Gamma_Z$ the mass and width of the $Z$-boson [115] and $N_c = 3$ the number of QCD colour charges.

The experimental resolution in the dimuon invariant mass is found to have an important effect on the core function, since it produces a significant shape variation in the mass region just above the $Z$-boson resonance and thus influences the lower end of the fit region in the $H \rightarrow \mu\mu$ search. To take this effect into account, the DY function described above is convoluted with a Gaussian function with mass dependent resolution derived from the simulation.

The core function is multiplied by the empirical component to obtain the final background parameterisation used in the fits to the $m_{\mu\mu}$ spectra. Two families of functions are studied for this empirical component: power law functions (“Power”) and the exponential of polynomials (“Epoly”), the definition of whose are summarised in Table 3.

<table>
<thead>
<tr>
<th>Function</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerN</td>
<td>$m_{\mu\mu}^{(a_0 + a_1 m_{\mu\mu} + a_2 m_{\mu\mu}^2 + \ldots + a_N m_{\mu\mu}^N)}$</td>
</tr>
<tr>
<td>EpolyN</td>
<td>$\exp(a_1 m_{\mu\mu} + a_2 m_{\mu\mu}^2 + \ldots + a_N m_{\mu\mu}^N)$</td>
</tr>
</tbody>
</table>

The criteria defined to select the background functions among those listed above and to determine the associated systematic uncertainty, referred to as the spurious signal (SS), are described in the following.

As first requirement, only functions able to fit the data sidebands, the fully simulated background samples and the fast DY simulation\(^2\) with a $\chi^2$ probability of the fit greater than 1% (for all these samples) are considered.

Among the functions that pass these criteria, a spurious signal test is performed in each BDT category using the fast DY simulation sample, as it is the sample with the highest statistical accuracy. The SS are defined as the measured signal yield when performing a signal plus background fit to a background MC template. They are determined not only for a signal mass of 125 GeV, but also varying $m_H$ between 120 and 130 GeV in steps of 1 GeV. Only the functions with the absolute value of the SS below 20% of the expected signal statistical error are considered. When applying this requirement, the MC statistical error is subtracted from the SS value. Among the functions that pass these requirements, the one with the smallest degrees of freedom is selected to ensure a higher sensitivity in the search. If multiple functions pass these

\(^2\) Unless mentioned otherwise, the fast DY simulation are reweighted using first or second order polynomial functions in $m_{\mu\mu}$ to the data sidebands.
requirements the one with the smallest SS is selected. The maximum absolute value of the SS in the mass range described above is considered as the background modelling uncertainty. This procedure is applied to each BDT category.

As an additional cross-check, the SS tests are also performed on the fast DY simulation after applying several theoretical and experimental uncertainties, such as variations of the QCD renormalisation and factorisation scales by factors of two up and down, alternative PDF sets and variations on the muon momentum resolution and scale within the experimental uncertainties. In these checks no significant increase of the SS values are found.

The list of the selected functions and the corresponding background systematic uncertainties expressed as the SS normalised to the expected signal statistical uncertainties are shown in Table 4. All the SS are considered as uncorrelated systematic uncertainties among the different BDT categories, since no values are found that are more than two standard deviations away from zero for a signal mass of 125 GeV after the above criteria. This considers the statistical accuracy of the fast DY simulation corresponding to about 700 times the data statistics.

Table 4: Selected empirical background functions in the different analysis categories together with the maximum values of the SS (in the 120–130 GeV mass range) normalised to the expected signal statistical error (δS) and to the SM predictions (S_{SM}) in %.

<table>
<thead>
<tr>
<th>Category</th>
<th>Empirical Function</th>
<th>max(SS/ δS)[%]</th>
<th>max(SS/ S_{SM})[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBF High</td>
<td>Power0</td>
<td>10.6</td>
<td>14.7</td>
</tr>
<tr>
<td>VBF Medium</td>
<td>Epoly2</td>
<td>0.51</td>
<td>1.3</td>
</tr>
<tr>
<td>VBF Low</td>
<td>Power1</td>
<td>3.6</td>
<td>7.5</td>
</tr>
<tr>
<td>2-jet High</td>
<td>Epoly2</td>
<td>8.7</td>
<td>16.3</td>
</tr>
<tr>
<td>2-jet Medium</td>
<td>Epoly4</td>
<td>1.2</td>
<td>3.9</td>
</tr>
<tr>
<td>2-jet Low</td>
<td>Epoly3</td>
<td>-8.2</td>
<td>-33.2</td>
</tr>
<tr>
<td>1-jet High</td>
<td>Power1</td>
<td>6.1</td>
<td>12.1</td>
</tr>
<tr>
<td>1-jet Medium</td>
<td>Epoly3</td>
<td>-8.1</td>
<td>-19.8</td>
</tr>
<tr>
<td>1-jet Low</td>
<td>Epoly3</td>
<td>-2.5</td>
<td>-5.8</td>
</tr>
<tr>
<td>0-jet High</td>
<td>Power1</td>
<td>14.6</td>
<td>26.5</td>
</tr>
<tr>
<td>0-jet Medium</td>
<td>Epoly3</td>
<td>-11.6</td>
<td>-39.0</td>
</tr>
<tr>
<td>0-jet Low</td>
<td>Epoly3</td>
<td>-18.5</td>
<td>-74.2</td>
</tr>
</tbody>
</table>

### 6.3 Other Systematic uncertainties

In addition to the systematic uncertainties on the signal and background modelling described above, also the uncertainty in the combined 2015-2018 integrated luminosity of 1.7% is considered. It is derived from the calibration of the luminosity scale using x–y beam-separation scans [116], following a methodology similar to that detailed in Ref. [117].
7 Results

The signal yield is obtained by a simultaneous binned maximum likelihood fit to the twelve $m_{\mu\mu}$ distributions of the BDT categories in the range 110–160 GeV. The bin size is set to 0.1 GeV. Confidence intervals are based on the profile-likelihood-ratio test statistics [118]. The systematic uncertainties listed in Section 6 are implemented in the fit as nuisance parameters constrained by additional Gaussian or log-Normal likelihood terms.

The twelve mass spectra together with the background and signal fit to data are shown in Figure 2. Inclusive spectra of the dimuon invariant mass after the signal plus background fit are presented in Figure 3. In Figure 3(b) the events are weighted by $\log(1+S/B)$, where $S$ and $B$ are signal and background yields derived from the fit to data in the 120–130 GeV window. The $\chi^2$ probabilities of the fits to the data in the twelve categories range between 20 and 94%.

The number of signal and background events estimated by the fit to data in a mass window of 120–130 GeV are shown in Table 5.

Table 5: Number of events observed in the $m_{\mu\mu} = 120–130$ GeV window in data, the number of signal events expected in the SM ($S_{SM}$), and events from signal ($S$) and background ($B$) as derived from the combined fit. In addition the observed number of signal events over square root of background events ($S/\sqrt{B}$) and the signal-to-background ratio ($S/B$) in % for each of the twelve BDT categories described in the text are displayed.

<table>
<thead>
<tr>
<th>Category</th>
<th>Data</th>
<th>$S_{SM}$</th>
<th>$S$</th>
<th>$B$</th>
<th>$S/\sqrt{B}$</th>
<th>$S/B$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBF High</td>
<td>40</td>
<td>4.5</td>
<td>2.3</td>
<td>34</td>
<td>0.39</td>
<td>6.6</td>
</tr>
<tr>
<td>VBF Medium</td>
<td>109</td>
<td>5.5</td>
<td>2.8</td>
<td>100</td>
<td>0.28</td>
<td>2.8</td>
</tr>
<tr>
<td>VBF Low</td>
<td>450</td>
<td>9.6</td>
<td>4.9</td>
<td>420</td>
<td>0.24</td>
<td>1.2</td>
</tr>
<tr>
<td>2-jet High</td>
<td>3400</td>
<td>38</td>
<td>19</td>
<td>3440</td>
<td>0.33</td>
<td>0.6</td>
</tr>
<tr>
<td>2-jet Medium</td>
<td>13938</td>
<td>70</td>
<td>35</td>
<td>13910</td>
<td>0.30</td>
<td>0.3</td>
</tr>
<tr>
<td>2-jet Low</td>
<td>40747</td>
<td>75</td>
<td>38</td>
<td>40860</td>
<td>0.19</td>
<td>0.1</td>
</tr>
<tr>
<td>1-jet High</td>
<td>2885</td>
<td>32</td>
<td>16</td>
<td>2830</td>
<td>0.31</td>
<td>0.6</td>
</tr>
<tr>
<td>1-jet Medium</td>
<td>24919</td>
<td>107</td>
<td>54</td>
<td>24890</td>
<td>0.35</td>
<td>0.2</td>
</tr>
<tr>
<td>1-jet Low</td>
<td>77482</td>
<td>134</td>
<td>68</td>
<td>77670</td>
<td>0.24</td>
<td>0.1</td>
</tr>
<tr>
<td>0-jet High</td>
<td>24777</td>
<td>85</td>
<td>43</td>
<td>24740</td>
<td>0.27</td>
<td>0.2</td>
</tr>
<tr>
<td>0-jet Medium</td>
<td>85281</td>
<td>155</td>
<td>79</td>
<td>85000</td>
<td>0.27</td>
<td>0.1</td>
</tr>
<tr>
<td>0-jet Low</td>
<td>180478</td>
<td>144</td>
<td>73</td>
<td>180000</td>
<td>0.17</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>

The best fit value of the signal strength parameter from the fit is $\mu = 0.5 \pm 0.7$, corresponding to an observed (expected) significance of $0.8\sigma$ ($1.5\sigma$) with respect to the hypothesis of no $H \rightarrow \mu\mu$ signal. The signal strength uncertainty is dominated by the statistical error on the data of about 0.7. The impact of the systematic uncertainties on the signal strength is found to be $+0.2_{-0.1}$, dominated by the signal theory uncertainties that account for 0.08, the signal experimental uncertainties that account for 0.07 and the spurious signal uncertainties that account for 0.06. A statistical test is performed to check the compatibility of the measured signal strengths among the 12 categories and found to be at the level of 33%.

An upper limit on $\mu$ is computed using a modified frequentist CL$_s$ method [118, 119]. The observed upper limit on $\mu$ at 95% CL is found to be 1.7, with an expected limit of 1.3 for the case of no $H \rightarrow \mu\mu$ signal and an expected limit of 2.2 for the case of a $H \rightarrow \mu\mu$ signal at SM strength. The corresponding branching
Figure 2: Signal plus background fits in the twelve BDT analysis categories. In the top panel the signal-plus-background model is shown (blue curve) overlaid to the data points. The signal component is shown separately (red line). In the bottom panel the difference between the data and the background function is shown with overlaid the fitted signal component (red curve).
ratio upper limit at 95% confidence level is $\text{Br}(H \rightarrow \mu\mu) < 3.8 \cdot 10^{-4}$, assuming the SM cross section for the Higgs boson production.

This result represents an improvement of about 50% in expected sensitivity compared with the previous ATLAS result [15]. Approximately half of this improvement comes from the increased integrated luminosity and half from refinements in the analysis techniques.

8 Conclusion

A search for the rare dimuon decay of the Higgs boson is performed using the full Run 2 data set of 139 fb$^{-1}$ collected with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 13$ TeV at the LHC. No significant excess is observed in data, and an upper limit on the signal strength at 1.7 is set at 95% CL, while 1.3 was expected for the case of no $H \rightarrow \mu\mu$ signal. The best fit value for the SM $H \rightarrow \mu\mu$ signal strength parameter is found to be $\mu = 0.5 \pm 0.7$. This represents an improvement of about 50% in sensitivity to the $H \rightarrow \mu\mu$ signal compared to the previous ATLAS result [15].

References


ATLAS Collaboration, *Observation of \(H \rightarrow \mu\mu \) decays and \(BR(h \rightarrow \tau\tau)\) does not equal \(m_\mu^2/m_\tau^2\)*, JHEP 05 (2013) 039, arXiv: 1302.3229 [hep-ph].


Appendix A

Figures 4 and 5 show the impact of the FSR recovery procedure on various $m_{\mu\mu}$ distributions. Figure 6 shows the invariant mass spectrum for data and full simulation MC (normalised to the number of data events) after preselection in the 76–160 GeV mass range. Figure 7 shows the $m_{\mu\mu}$ data distributions for the 12 BDT categories with overlaid the fit results, including the different background and signal components. Figure 8 shows the $m_{\mu\mu}$ distributions from fast DY simulation reweighted to data sidebands for the 12 BDT categories with overlaid the fit results, including the different background and signal components. Figure 9 shows the signal strength per category and combined obtained from the combined fit on observed data and from Asimov data. Figure 10 shows comparison between the data and the fast DY simulation of the $m_{\mu\mu}$ spectrum in the 110-160 GeV mass regions. Figure 11 shows the BDT distribution for the different n-jet categories for data for different mass regions, as well as signal ggF and VBF MC. Figure 12 shows the signal and background composition for the twelve BDT categories. Figure 13 shows the signal mass parameterisation for two BDT categories. Figure 14 shows the BDT training variables for $n_j = 0$ and $n_j = 1$ categories.

Figure 4: Invariant mass of $\mu\mu(\gamma)$ final states for events with a reconstructed FSR photon candidate (left) and for all ggF signal events (right). The black and blue histograms represent the distributions before and after the FSR recovery, respectively. Histograms are scaled to 139 fb$^{-1}$.
Figure 5: The invariant mass of $\mu\mu(\gamma)$ final states for events with a reconstructed FSR photon candidate in the region around the Z-boson resonance. The black and blue histograms represent the distributions before and after the FSR recovery for $Z \rightarrow \mu\mu$ MC events scaled to $139 \text{ fb}^{-1}$. The black and blue circles represent data before and after the FSR recovery respectively.

Figure 6: Inclusive dimuon invariant mass $m_{\mu\mu}$ in the range 76–160 GeV after the analysis preselection. Data in points are compared to a full set of fully simulated background processes and the total background prediction is scaled to the integrated data yield. The shaded region in the bottom panel shows the impact of the systematic uncertainty on the muon momentum scale and resolution. The $H \rightarrow \mu\mu$ signal is shown for the $ggF$, VBF and VH processes in open lines, normalised to one hundred times the SM prediction for visibility.
Figure 7: Signal plus background fits in the twelve BDT analysis categories. In the top panel the signal plus background function is shown (blue curve) overlaid to the data points, together with the signal (red line) and core function (green dashed line) components. Also given are the values of $\chi^2$ per degree of freedom of the fit and the corresponding $\chi^2$-probability in percent (labelled with $\chi^2$ and $p$). In the middle panel the ratio between the data and the signal plus empirical function both divided by the core component are shown. In the bottom panel the difference between the data and the background function is shown with overlaid the fitted signal (red curve).
Figure 8: Illustration of the background-modeling procedure using the high-statistics fast DY simulation (points) with an equivalent luminosity of about 100 ab$^{-1}$. The top panel shows the dimuon invariant mass spectra in the various BDT categories overlaid with background-only fit (blue curve) and core function (green dashed line). Also given are the values of $\chi^2$ per degree of freedom of the background-only fit and the corresponding $\chi^2$-probability in percent (labelled with $\chi^2$ and $p$). The red line in the top and bottom panels represents the signal PDF normalised to the predicted cross-section, for the bottom panel it is divided by the background PDF. The dashed red horizontal lines in the bottom panel give the expected $S/B$ in range of $m_{\mu\mu} = 120–130$ GeV. The middle panels show the ratio between the fast DY simulation and the signal plus empirical function both divided by the core component are shown. In the bottom panel the ratio between the fast DY simulation and the background function is shown.
Figure 9: Expected (left) and observed (right) signal strengths in each category and combined for all categories.

Figure 10: Normalised di-muon invariant mass spectra in four selected BDT categories, comparing data with non-DY components subtracted (points) to ATLAS full DY simulation with Sherpa (blue line) and the fast DY simulation (red line).
Figure 11: The multivariate classifiers used in the analysis for (a) 0-jet events, (b) 1-jet events, (c)+(d) 2-jet events. In each panel the normalised distributions are shown for signal from VBF (orange line) and ggF production (green line) as well as the data sidebands used in the training (black line). In addition, the observed data in the region $m_{\mu\mu} = 120$–130 GeV are overlaid.
Figure 12: Summary of signal and background composition in the twelve analysis categories in the mass region $m_{\mu\mu} = 120–130$ GeV. The top panel shows the number of background events ($B$, multiplied by a factor of $10^{-5}$), the expected signal-to-background ratio for a SM $H \rightarrow \mu\mu$ signal ($S/B$, multiplied by 10), as well as the counting-significance $S/\sqrt{B}$. The middle panel shows the contributions of different Higgs-boson production modes to the total signal in each category. The bottom panel shows the background composition from Drell–Yan, Diboson and top-quark processes.
Figure 13: Dimuon invariant mass spectra of signal $H \rightarrow \mu\mu$ events in two specific BDT categories of the analysis. For both categories the distribution from the signal simulation is shown in points and the parametric signal model fitted to the distributions is shown as line. The central values $m_{CB}^0$ as well as the width $\sigma_{CB}$ of the signal model Crystal-Ball functions are also shown.
Figure 14: The distributions of the three variables used in the BDT for $n_j = 0$ events (top row) and the six variables used in the BDT for $n_j = 1$ events (middle and lower row) for $H \rightarrow \mu\mu$ signal MC (red), data sideband (black) and background MC sideband and center (defined as 120-130 GeV mass range). The background MC includes DY from full simulation, $t\bar{t}$, diboson, single top-quark events.