Evidence for Higgs boson decays to a low-mass dilepton system and a photon in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

DOI
10.1016/j.physletb.2021.136412

Publication date
2021

Document Version
Final published version

Published in
Physics Letters B

License
CC BY

Citation for published version (APA):

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

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)

Download date: 16 Oct 2023
Evidence for Higgs boson decays to a low-mass dilepton system and a photon in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector

The ATLAS Collaboration *

A R T I C L E   I N F O

Article history:
Received 19 March 2021
Received in revised form 11 May 2021
Accepted 25 May 2021
Available online 31 May 2021
Editor: M. Dosier

A B S T R A C T

A search for the Higgs boson decaying into a photon and a pair of electrons or muons with an invariant mass $m_{\ell\ell}<30$ GeV is presented. The analysis is performed using 139 fb$^{-1}$ of proton–proton collision data produced by the LHC at a centre-of-mass energy of 13 TeV and collected by the ATLAS experiment. Evidence for the $H\rightarrow\ell\ell\gamma$ process is found with a significance of 3.2 over the background-only hypothesis, compared to an expected significance of 2.1 for the Standard Model prediction. 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 $\mu=1.5\pm0.5$. The Higgs boson production cross-section times the $H\rightarrow\ell\ell\gamma$ branching ratio for $m_{\ell\ell}<30$ GeV is determined to be $8.7^{+4.1}_{-3.7}$ fb.

© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). Funded by SCOAP$^3$.

1. Introduction

In July 2012, the ATLAS and CMS Collaborations at the CERN Large Hadron Collider (LHC) announced the discovery of a new particle with a mass of approximately 125 GeV [1,2]. The observed properties of the particle, such as its couplings to Standard Model (SM) elementary particles, its spin and its parity, are so far consistent with the predictions for the SM Higgs boson [3–7].

Measurements of rare decays of the Higgs boson, such as $H\rightarrow\ell\ell\gamma$ where $\ell$ is an electron or muon, can probe coupling modifications introduced by possible extensions to the SM [8]. In addition, such three-body Higgs boson decays can be used to probe CP-violation in the Higgs sector [9,10].

Multiple processes contribute to the $H\rightarrow\ell\ell\gamma$ decay: Dalitz decays involving a Z boson or a virtual photon ($\gamma^*$) (Fig. 1(a–c)), as well as the decay of the Higgs boson to two leptons and a photon from final-state radiation (FSR) (Fig. 1(d)). Their respective fractions depend on the invariant mass of the dilepton pair, $m_{\ell\ell}$. In this analysis only low-mass dilepton pairs with $m_{\ell\ell}<30$ GeV are considered. This region is completely dominated by the decay through $\gamma^*$ [8,11,12]. The contributions of the other processes and interferences are negligible.

Based on a data sample of proton–proton (pp) collisions at $\sqrt{s}=13$ TeV with an integrated luminosity of 35.9 fb$^{-1}$, the CMS Collaboration reported a 95% CL upper limit on the production cross-section times branching ratio for the low-mass $H\rightarrow\mu\mu\gamma$ process of 4.0 times the SM prediction [13]. In addition, both the ATLAS and CMS Collaborations carried out searches at $\sqrt{s}=13$ TeV for the closely related $H\rightarrow Z\gamma$ process [13,14]. The CMS Collaboration also searched for the low-mass $H\rightarrow\ell\ell\gamma$ process in the dimuon and dielectron channels in pp collisions at $\sqrt{s}=8$ TeV [15].

This paper describes a search for $H\rightarrow ee\gamma$ and $H\rightarrow \mu\mu\gamma$ decays with $m_{\ell\ell}<30$ GeV. When the invariant mass of the two electrons is low and the transverse momentum of the dielectron system is high, their electromagnetic showers can overlap in the calorimeter. Therefore, the search for $ee\gamma$ final states requires the development of dedicated electron trigger and identification algorithms.

The search uses pp collision data at $\sqrt{s}=13$ TeV recorded with the ATLAS detector during Run 2 of the LHC between 2015 and 2018, corresponding to a total integrated luminosity of 139 fb$^{-1}$. The sensitivity of the search is enhanced by dividing the selected events into mutually exclusive categories, according to the event topology and lepton flavour. The dominant background is the irreducible non-resonant production of $\ell\ell\gamma$. After event categorisation, the signal yield is extracted by a simultaneous fit of parametric functions to the reconstructed $\ell\ell\gamma$ invariant mass ($m_{\ell\ell\gamma}$) distributions in all categories.

2. ATLAS detector

The ATLAS detector [16] covers nearly the entire solid angle around the collision point.1 It consists of an inner tracking de-

---

1 The ATLAS experiment 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...
tector (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic (EM) and hadron calorimeters, and a muon spectrometer (MS). The ID covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. The silicon pixel detector provides up to four measurements per track. The insertable B-layer (IBL) [17,18] constitutes the innermost layer at a radius of 33.3 mm. It surrounds the beam pipe, which has an inner radius of 23.5 mm. Lead/liquid-argon (LAr) sampling calorimeters provide EM energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap and forward regions are instrumented with LAr calorimeters for both the EM and hadronic energy measurements up to $|\eta| = 4.9$. For $|\eta| < 2.5$, the EM calorimeter is divided into three longitudinal layers, which are finely segmented in $\eta$ and $\phi$, particularly in the first layer. This segmentation allows the measurement of the lateral and longitudinal shower profile, and the calculation of shower shapes [19] used for particle identification and background rejection. The longitudinal segmentation of the EM calorimeter is also exploited to calibrate the energy response of electron and photon candidates [19]. The 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.

A two-level trigger system [20] was used during the $\sqrt{s} = 13$ TeV data-taking period. The first-level trigger (L1) is implemented in hardware and uses a subset of the detector information. This is followed by a software-based high-level trigger which runs algorithms similar to those in the offline reconstruction software, reducing the event rate to approximately 1 kHz from the maximum L1 rate of 100 kHz.

3. Data and simulated event samples

The analysed $pp$ collision data correspond to the full recorded LHC Run-2 data set at $\sqrt{s} = 13$ TeV with an integrated luminosity of 139 fb$^{-1}$ after data quality requirements [21]. The events were collected with a combination of single-lepton, dilepton, diphoton, and lepton+photon triggers. Standard electron triggers, which require narrow, isolated EM energy clusters, are efficient if the two electrons in the event have a very small angular separation $\Delta R_{ee}$, as the two electrons produce a cluster that is similar to that of a single electron. For $\Delta R_{ee} > 0.1$, two separate EM clusters are typically produced. For the 2017 and 2018 data-taking periods, a dedicated trigger was introduced which requires at least one photon with $p_T > 35$ GeV and at least one EM cluster with $p_T > 25$ GeV matched to at least one ID track. The EM cluster must have low hadronic leakage, but no requirement is imposed on the width of the shower, which allows this new trigger to recover a significant fraction of the signal with $0.025 < \Delta R_{ee} < 0.1$. More details are provided in Ref. [22]. The overall trigger efficiencies for the $H \rightarrow ee$ and $H \rightarrow \mu\mu\gamma$ processes are 98% and 96%, respectively (97% combined), relative to the event selection discussed in Section 5. The uncertainty in the trigger efficiency is negligible.

Samples of simulated Monte Carlo (MC) events are crucial for the optimisation of the search strategy and the modelling of the signal and background processes. Simulated $H \rightarrow \gamma^*\gamma \rightarrow \ell\ell\gamma$ signal and $H \rightarrow \gamma\gamma$ background events are used to parameterise these processes, and simulated non-Higgs $\ell\ell\gamma$ events are used to choose analytic functional forms to describe the non-resonant background. The generated MC events, unless stated otherwise, were processed with the full ATLAS detector simulation [23] based on GEANT4 [24]. The effect of multiple $pp$ interactions in the same and neighbouring bunch crossings (pile-up) was included by over-laying minimum-bias events simulated with Pythia 8.186 [25] using the NNPDF2.3LO parton distribution function (PDF) set [26] and the A5 set [27] of tuned parameters. The MC events were weighted to reproduce the distribution of the number of interactions per bunch crossing observed in the data.

Higgs boson production in the gluon–gluon fusion (ggF) and vector-boson fusion (VBF) production modes, as well as in quark-initiated associated production with a $W$ or $Z$ boson ($q\bar{q}/gg \rightarrow VH$) or with two top quarks ($t\bar{t}H$) was modelled with the PowHEG-Box v2 MC event generator [28–32]. PowHEG-Box v2 was interfaced with Pythia 8 [25] to simulate the $H \rightarrow \gamma^*\gamma \rightarrow \ell\ell\gamma$ and $H \rightarrow \gamma\gamma$ decays. Pythia also provides parton showering, hadronisation, and the underlying event. The PDF4LHC15 PDF set [33] was used, except for $t\bar{t}H$, where the NNPDF3.0lo PDF set [26] was used. For the ggF process, the PowHEG NNLOPS program [34,35] achieves next-to-next-to-leading-order (NNLO) accuracy in QCD for inclusive observables after reweighting the Higgs boson rapidity spectrum [36]. The simulation reaches next-to-leading-order (NLO) accuracy in QCD for the VBF, VH, and $t\bar{t}H$ processes. For VH, the MiNLO technique [37–39] was applied. No simulated samples are available for gluon-initiated associated production with a $Z$ boson ($gg \rightarrow ZH$) and associated production with two $b$-quarks ($bbH$). Their minor contribution, $1\%$ of the expected Higgs boson production, is modelled using the acceptances from the $q\bar{q}/gg \rightarrow ZH$ and $ggF$ samples, respectively.

Alternative $H \rightarrow \gamma\gamma$ samples are considered for the ggF and VBF processes, where HERWIG 7 [40] was used instead of Pythia 8 to provide parton showering. Weights were calculated by comparing the generator-level Higgs boson transverse momentum and jet distributions in these samples with the distributions obtained from

---

Fig. 1. Representative Feynman diagrams of the $H \rightarrow \ell\ell\gamma$ process.
the nominal $H \to gg$ samples. The weights are applied to $H \to \gamma\gamma$ with VBF and VBF samples to estimate the effect of varying the parton shower and underlying event.

The mass of the Higgs boson was set in the simulation to $m_H = 125$ GeV and its width to $\Gamma_H = 4.1$ MeV [41]. The Higgs mass peak position in the simulation was corrected to account for the small difference relative to the measured value of 125.09 GeV [42] (see Section 6); otherwise the effect on the kinematics is neglected.

The signal and $H \to \gamma\gamma$ samples were normalised with the best available theoretical calculations of the corresponding SM production-cross-sections at $m_H = 125.09$ GeV, which are available at next-to-next-to-next-to-leading order (N3LO) in QCD with NLO electroweak (EW) corrections for ggF [41,43–54], at approximate-NLLO in QCD with NLO EW corrections for VBF [41,55–57], at NNLO in QCD with NLO EW corrections for $qg/gg \to VH$ [41,58–65] and at NLO in QCD with NLO electroweak corrections for $t\bar{t}H$ [41,66–69]. The accuracy of the calculation is NNLO in QCD for $bbH$ [70–72], and NLO in QCD with next-to-leading-logarithm (NLL) corrections for $gg \to ZH$ [41,61].

There are multiple calculations of the $H \to \ell\ell$ branching ratios for different slices of phase space [8,11,12,73–76]. The $H \to \ell\ell$ branching ratios used in this analysis are $B(H \to e\mu) = 7.20 \times 10^{-5}$ and $B(H \to \mu\mu) = 3.42 \times 10^{-3}$, as estimated with PyTHIA 8 for $m_H < 30$ GeV. Their sum corresponds to ~5% of the $H \to \gamma\gamma$ branching ratio. When extrapolated to a common phase space, the PyTHIA 8 estimate agrees with the predictions of Refs. [74,76] within a relative difference of 3%. Currently, no theoretical uncertainty calculation exists for the low-$m_H$ $H \to \ell\ell$ branching ratio. Therefore, the theoretical uncertainties in the $H \to \gamma\gamma$ and $H \to Z\gamma$ branching ratios are considered [12], and the larger of the two uncertainties is used, which corresponds to the 5.8% relative uncertainty in the $H \to Z\gamma$ branching ratio [41].

The background originates mainly from non-resonant $\ell\ell$ production. Events were simulated with the Sherpa 2.2.8 [77] generator based on LO matrix elements for $\ell\ell$ production with up to three additional partons and using the NNPDF3.0 PDF set [78]. The Sherpa simulation includes parton shower, fragmentation and underlying-event modelling. As the statistical uncertainties in the simulated $\ell\ell$ background samples are a limiting factor when studying the background modelling at the level required by the small signal-to-background ratio, a procedure was developed to generate significantly larger samples. These are based on events generated using the Sherpa 2.2.8 configuration described above with object efficiencies approximated by parameterisations rather than using the full ATLAS detector simulation and reconstruction software. The parameterisations, extracted from fully simulated MC samples, reproduce the reconstruction and selection efficiencies of detector-level objects via event weighting. Comparisons with a sample that underwent the full ATLAS detector simulation, and whose effective luminosity is greater than the data luminosity, show good agreement within the statistical uncertainties.

4. Object selection

Events are required to have a collision vertex associated with at least two tracks with transverse momentum $p_T > 0.5$ GeV each. In the case of multiple vertices, the vertex with the largest $\sum p_T^2$ of the associated tracks is considered to be the primary vertex.

Muon candidates are obtained by matching high-quality tracks in the MS and ID. Standalone MS tracks are used to extend the muon reconstruction beyond the ID acceptance to $2.5 < |\eta| < 2.7$. Muon candidates are required to satisfy the medium identification criteria [79,80], be within $|\eta| < 2.7$ and have $p_T > 3$ GeV. Muon candidates with an associated ID track must be matched to the primary vertex by having a longitudinal impact parameter $\Delta_d$ that satisfies $|\Delta_d \sin \theta| < 0.5$ mm, where $\theta$ is the polar angle of the track. The significance of the transverse impact parameter $d_0$ calculated relative to the measured beam-line position is required to be $|d_0|/\sigma_d < 3$, where $\sigma_d$ is the uncertainty in $d_0$. A subset of the muon candidates are also required to be isolated from additional activity in the tracking detector and in the calorimeters, using a loose isolation selection [79,80]. The efficiency of the muon reconstruction and identification, as well as the momentum calibration, including the associated systematic uncertainties, is estimated as described in Refs. [79,80].

Photon and electron candidates are reconstructed from energy clusters in the calorimeters which are formed using an algorithm based on dynamical, topological cell-clustering [19]. Candidates in the transition region between the barrel and endcap EM calorimeters, $1.37 < |\eta| < 1.52$, are excluded. The performance of the electron and photon reconstruction, including the associated systematic uncertainties, is studied in Ref. [19].

Photon candidates can be either unconverted or converted. In the latter case, a track or conversion vertex is matched to the EM cluster. Photon candidates are required to satisfy $p_T > 20$ GeV and $|\eta| < 0.37$, and pass the tight identification requirements [19,81] as well as a loose isolation selection [19].

The energy of the EM clusters associated with the photon and electron candidates is corrected in successive steps using a combination of simulation-based and data-driven correction factors [19]. The simulation-based calibration regression is optimised separately for electrons, unconverted photons and converted photons. The resolution of the energy response is corrected in the simulation to match the resolution observed in data using $Z \to ee$ events, by smearing the electron energy such that the width of the simulated $Z$ boson peak matches the width observed in data.

Because of the event kinematics of the signal process, it is common for the energy deposits of the two electrons in the EM calorimeter to be reconstructed as a single cluster. Therefore, two types of electron candidates are defined, each with its own selection criteria: one in which a topological cluster of energy deposits is associated with one selected ID track (resolved electron [82]), representing a single electron, and one in which a topological cluster is associated with two selected ID tracks (merged $ee$), representing a merged electron pair. Each ID track considered must satisfy $|\Delta_{Z_0} \cdot \sin \theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} < 5$. Resolved electron candidates must satisfy the medium likelihood–based identification criteria [19,82], have $p_T > 4.5$ GeV and be within $|\eta| < 2.47$. A subset of the resolved electron candidates are also required to pass a loose isolation requirement [19].

ID tracks considered for merged electrons must have opposite charge, $p_T > 5$ GeV, $|\eta| < 2.5$, and at least seven hits in the pixel and microstrip detectors combined. To suppress backgrounds from converted photons, tracks must also have a hit in the innermost pixel layer, and merged-electron candidates are rejected if they match a conversion vertex with a radius larger than 20 mm whose momentum agrees better with the cluster energy than the momentum of the track that geometrically matches the cluster best. Because the kinematic behaviour of merged electrons in the calorimeter most closely resembles a photon converting in the material close to the interaction vertex, the energy of the merged–$ee$ object is calibrated as a converted photon with a conversion radius set to $r_{conv} = 30$ mm calculated relative to the measured beam-line position. As presented in Fig. 2(a), this treatment is found to induce minimal bias in the dielectron energy measurement. To cover remaining differences between the simulated detector response to converted photons and merged–$ee$ objects, the difference between the squares of the energy resolutions for converted photons and merged–$ee$ objects in the MC simulation is treated as the square of an additional energy resolution uncertainty. The four-momentum of the merged–$ee$ candidate is constructed using the calibrated en-
ergy and the direction and invariant mass obtained from the vertex reconstructed from the two electron-track candidates [83].

Merged-ee candidates are required to have $|\eta| < 2.37$ (excluding $1.37 < |\eta| < 1.52$) and $p_T > 20$ GeV, and satisfy dedicated identification requirements, as the standard electron criteria have a low efficiency for objects with closely spaced energy deposits or broader EM showers. For the merged-ee identification, a multivariate discriminator is trained to separate the $\gamma^* \rightarrow ee$ signal objects from jets or single electrons. The input variables for the training include shower shape variables [82], the information provided by the transition radiation tracker [84], and the kinematic information from the cluster and ID tracks. Merged-ee candidates are also required to pass a tight isolation requirement [19]. A combined efficiency of $\sim50\%$ for the merged-ee identification and isolation is achieved for $H \rightarrow \gamma^* \gamma \rightarrow ee\gamma$ events. Since photons with a relatively small conversion radius offer a signature similar to that of merged-ee objects, the merged-ee identification and isolation efficiency is measured in data using a tag-and-probe method with FSR photons from Z boson radiative decays ($Z \rightarrow \ell\ell\gamma$). Candidate $Z \rightarrow \ell\ell\gamma$ events are selected similarly to those in Ref. [81].

Only two-track converted photons with $r_{\text{con}} < 160$ mm, corresponding to conversions inside the silicon pixel detector volume, are considered. The $Z \rightarrow \ell\ell\gamma$ and background yields are estimated from a fit to the $m_{\ell\ell\gamma}$ distribution in data. The extracted efficiencies are compared with the efficiencies estimated from simulated $Z \rightarrow \ell\ell\gamma$ events as shown in Fig. 2(b) for photons with $|\eta| < 0.8$. The resulting $p_T$- and $\eta$-dependent data/simulation scale factors are between 0.9 and 1.1 and are used to correct the identification and isolation efficiencies of the simulated $H \rightarrow \gamma^* \gamma \rightarrow ee\gamma$ events. The statistical uncertainties of the scale factors are taken into account. In addition, a systematic uncertainty is assessed for the background modelling by varying the selection criteria of the $m_{\ell\ell\gamma}$ background template. The total uncertainty reaches 2% for $20 < p_T < 30$ GeV and 9% for $p_T > 50$ GeV. Fig. 2(b) also shows a comparison of the extracted efficiencies with efficiencies in simulated $H \rightarrow \gamma^* \gamma \rightarrow ee\gamma$ events where an additional generator-level requirement of $|\Delta p_T| < 0.003$ is used. This requirement is applied in order to better match the signal signature to the converted photon signature in the detector; approximately 70% of the merged-ee objects in the signal sample pass it. The efficiencies of converted photons and the merged-ee objects agree within 10% for $p_T < 30$ GeV and within 5% for $p_T > 30$ GeV in the entire $\eta$ range.

Jets are reconstructed from topological clusters [85] using the anti-$k_T$ algorithm [86] with a radius parameter of 0.4 and are required to have $p_T > 20$ GeV and $|\eta| < 4.4$. Jets produced in pile-up interactions are suppressed by requiring that those with $p_T < 60$ GeV and $|\eta| < 2.4$ pass a selection based on a jet vertex tagging algorithm [87].

5. Event selection

Candidate $H \rightarrow \gamma^* \gamma \rightarrow \ell\ell\gamma$ events must have at least one reconstructed photon, and at least one opposite-charge, same-flavour pair of leptons (muons or resolved electrons) or one merged-ee object. One of the muons (resolved electrons) in the lepton pair must have $p_T > 11$ (13) GeV, to match the $p_T$ thresholds used in the dilepton triggers. As discussed above, the merged-ee $p_T$ is required to be larger than 20 GeV. If the leading photon overlaps with one of the EM clusters (resolved electrons or merged-ee) forming the $\gamma^*$ candidate within $\Delta R < 0.02$, the photon is discarded and the next-highest-$p_T$ photon is considered. No isolation requirement is applied to the subleading lepton (muon or resolved electron) as it is within the isolation cone of the leading lepton in the majority of events. Additionally, if the subleading lepton (muon or electron) falls within the isolation cone of the leading lepton, it is not included in the calculation of the isolation variable.

Muon pairs are given the highest priority if there are lepton pairs of different flavours in one event. If no opposite-charge muon pair satisfying the requirements above is found, the electron-pair candidates are considered. The resolved electron pair or merged-ee object with the highest vector-sum of the $p_T$ of the associated ID tracks is selected.

In order to suppress events arising from FSR processes, events are rejected if the photon is within a $\Delta R = 0.4$ cone around either of the selected leptons. If the axis of a jet is within $\Delta R = 0.4$ of the photon or a muon, or within $\Delta R = 0.2$ of an electron (merged or resolved), the jet is discarded. However, if the electron–jet angular separation is $0.2 < \Delta R < 0.4$, the electron and therefore the event is discarded.

To suppress events involving $Z$ boson decays, the dilepton invariant mass must satisfy $m_{\ell\ell} < 30$ GeV. To remove $J/\psi \rightarrow \ell\ell$ production, events with a dimuon (dielectron) invariant mass in
the range $2.9 < m_{\ell\ell} < 3.3$ GeV ($2.5 < m_{\text{ee}} < 3.5$ GeV) are excluded. Similarly, events with a di-muon (dielectron) invariant mass in the range $9.1 < m_{\mu\mu} < 10.6$ GeV ($8.0 < m_{\text{ee}} < 11.0$ GeV) are rejected to avoid $\eta(nS) \to \ell\ell$ contamination.

The invariant mass of the $\ell\ell\gamma$ system is required to satisfy $110 < m_{\ell\ell\gamma} < 160$ GeV. Additionally, both the photon and dilepton momenta must satisfy $p_T > 0.3 m_{\ell\ell\gamma}$.

The selected events are classified into mutually exclusive categories, depending on the lepton types and event topologies. The VBF-enriched categories, which have the best expected signal-to-background ratio, but the lowest event count, are defined as follows. Events must contain at least two jets with $p_T > 25$ GeV. If the leading or subleading jet is a forward jet (defined as a jet with $|\eta| > 2.5$), it is required to have $p_T > 30$ GeV to suppress jets originating from pile-up. In addition, the invariant mass of the two leading jets, $m_{jj}$, must be greater than 500 GeV, and the pseudo-rapidity separation between the two leading jets, $|\Delta\eta_{jj}|$, greater than 2.7. The quantity $|\eta_{\ell\ell}\gamma| = 0.5(|\eta_{jj} + |\eta_{\ell\ell}|)$ [88] is required to be less than 2.0, where $|\eta_{\ell\ell}\gamma| > 2.0$ is the pseudorapidity of the $\ell\ell\gamma$ system and $\eta_{\ell\ell}$ is the pseudorapidity of the leading (subleading) jet. The selected leptons and the photon must be separated from the two jets by $\Delta R > 1.5$. Additionally, the azimuthal separation between the $\ell\ell\gamma$ system and the system formed by the two jets must be greater than 2.8. These jet variables are expected to have a different shape for the VBF signal and the background from QCD processes.

The $p_T$ is defined as the component of the transverse momentum of the $\ell\ell\gamma$ system that is perpendicular to the difference of the three-momenta of the dilepton system and the photon candidate $p_T = |p_T| \times t$, where $t = (p_T^1 - p_T^2) / |p_T^1 - p_T^2|$. This quantity is strongly correlated with the transverse momentum of the $\ell\ell\gamma$ system, but has better experimental resolution [89,90]. Events failing the VBF-enriched selection, but having $p_T > 100$ GeV, are assigned to the high-$p_T$ category, which has a lower expected signal-to-background ratio than the VBF-enriched category, but is expected to have more events. The high-$p_T$ category is expected to have an increased fraction of VBF and VH events as these production modes lead on average to higher Higgs boson $p_T$ than ggF production.

The vast majority of selected events do not fall into the VBF-enriched or high-$p_T$ categories discussed above, and are placed into the low-$p_T$ category. The full list of all categories considered in the analysis, together with the expected yields for a 125.09 GeV Higgs boson decaying into $\ell\ell\gamma$, is shown in Table 1. The table also summarises the observed number of events in data in the $m_{\ell\ell\gamma}$ mass range of 110–160 GeV.

6. Signal and background modelling

To extract the observed signal yield, parametric functions are chosen to model the $\ell\ell\gamma$ invariant mass distributions of signal and background in each analysis category. A combined model is built from these functions and fit to the selected data in the $m_{\ell\ell\gamma}$ range of 110–160 GeV, simultaneously in all categories.

The signal model, including its parameters, is obtained by fitting a double-sided Crystal Ball function (DSCB) [91,92] to the $m_{\ell\ell\gamma}$ distribution obtained from the $H \to \gamma\gamma \to \ell\ell\gamma$ samples described in Section 3 after applying the event selection and categorisation from Section 5. The DSCB function features a Gaussian core and asymmetric power-law tails. A shift of $+0.09$ GeV is applied to the mean of the Gaussian function to account for the fact that the sample assumes a Higgs boson mass of 125 GeV. The effective signal mass resolution, defined as half the width containing 68% of the signal events, depends on the category and lies between 1.6 GeV (ee-merged high-$p_T$ category) and 2.2 GeV (ee-resolved low-$p_T$ category).

The $H \to \gamma\gamma$ process, which contributes as background to the electron categories through converted photons, is modelled using the same functions and parameters as the signal in the respective categories, and is normalised to the predicted SM yield. The parameterisations are compatible with the statistically limited distributions obtained from simulated $H \to \gamma\gamma$ events. This background contribution is relatively small (<2.5% and <7% of the expected $H \to \gamma\gamma \to \ell\ell\gamma$ signal in the ee-merged channel and ee-merged channel, respectively) and is taken into account in the fitting procedure.

The non-resonant background is also estimated with parametric functions. The background normalisation and function parameters are allowed to float in the fit to data. The chosen functional forms are based on background templates that are built taking into account the contributions of different processes to the background. The function choice is performed separately for each category. The dominant part of the background originates from the non-resonant $\ell\ell\gamma$ process. There is also a smaller background from events with misidentified photons, electrons, or muons. To estimate the fraction of events with a misidentified photon, a control region is defined using the signal selection, but dropping the photon isolation requirement and using it as a discriminating variable in a template fit. A background template enriched in events with misidentified photons is built by inverting the photon identification selection, with the prompt-photon contamination removed by subtracting its distribution found in simulated events. The template normalisation is obtained in the background-dominated sideband of the photon isolation distribution for each category separately. In the signal region, about 10% of all selected events have a misidentified photon, independently of the category. The fraction of events with a misidentified electron or muon that is in fact a hadronic jet, a lepton from heavy-flavour decays or an electron from a photon conversion, is estimated in a similar manner, using a control region in which the isolation selection is dropped for the softer lepton. The estimated fractions of events with misidentified leptons are 4% in the $\mu\mu$ low-$p_T$ category, 2% in the ee-merged low-$p_T$ category, and 30% in the ee-resolved low-$p_T$ category. Events in the last category are separated into two groups depending on the angular distance between the electrons, as two populations with different misidentified-electron rates and mass distributions were found. The numbers in the other categories are extracted as well, but suffer from fairly large statistical uncertainties.

The invariant mass template for the non-resonant $\ell\ell\gamma$ background is built from the simulated events described in Section 3. The invariant mass templates for events with misidentified ob-
objects are obtained from background-dominated control regions, and scaled by the yields derived above. Reasonable agreement is observed between the templates containing the sum of all backgrounds and the sidebands of the $m_{\ell\ell}$ distributions in data ($105-120$ GeV and $130-160$ GeV).

The choice of fit function for the non-resonant background is made in each category using signal-plus-background fits to the constructed background-only template by measuring the bias associated with each function, expressed as the number of fitted signal events, and choosing the function with the smallest number of degrees of freedom that satisfies the bias criteria described below.

The functional forms used to model the background are selected from the following: exponential ($e^{\alpha m_{\ell\ell}}$), exponential of a second-order polynomial ($e^{\alpha m_{\ell\ell}+\beta m_{\ell\ell}^2}$), and a power-law function ($m_{\ell\ell}^{\alpha}$), where $\alpha$ and $\beta$ are free parameters. Signal hypotheses with $m_H$ ranging from 121 GeV to 129 GeV are tested in steps of 1 GeV, and the fit bias is evaluated as the absolute maximum bias over this range (referred to as the 'spurious signal' in the following), similar to what is done in Ref. [1]. A function passes the test if the spurious signal is less than 10% of the expected number of $H \rightarrow \gamma^*\gamma \rightarrow \ell\ell\gamma$ events or less than 20% of the statistical uncertainty of the fitted spurious signal due to the expected number of background events. To account for statistical fluctuations in the background template, the criteria are relaxed by the statistical uncertainty due to the template [93]. Furthermore, in a background-only fit of the template in the mass range $110 < m_{\ell\ell} < 160$ GeV, the function must pass a $\chi^2$ test with a probability larger than 1%. As described in Ref. [93], an additional check is performed by fitting both the chosen function and a function belonging to the same family with one more degree of freedom to the sidebands of the $m_{\ell\ell}$ distribution in data, to ensure the data do not prefer the more complex function. Following the procedure described above, the power-law function is selected for all categories, except for the $\mu\mu$ low-$p_T$, and ee-merged low-$p_T$ categories, which are best described by the second-order exponential polynomial, and the ee-resolved VBF-enriched category, for which the exponential function is chosen.

An unbinned extended likelihood function is formed from the product of each category's parameterised signal-plus-background probability density function. Systematic uncertainties are considered in the form of nuisance parameters with Gaussian or log-normal constraints. They are correlated across all categories, except for the spurious-signal uncertainties. The latter are implemented for each category as a signal-like component with a yield parameter that is constrained by a Gaussian function, centred at zero, with a width corresponding to the estimated spurious signal. The Higgs boson mass is set to $125.09 \pm 0.24$ GeV [42].

The parameter of interest is the signal strength $\mu$, which is defined as the ratio of the measured signal yield to the SM expectation, taking into account the uncertainties on the latter. The corresponding profile likelihood ratio is maximised to extract the best-fit $\mu$ [94]. A possible excess over the background-only hypothesis is quantified by a $p$-value using the profile likelihood ratio, evaluated for a vanishing $H \rightarrow \ell\ell\gamma$ branching ratio, as a test statistic. For all results, the asymptotic approximation is used [94]. Cross-checks of the best-fit $\mu$ and the $p$-value extracted with pseudo-experiments show good agreement with those obtained using the asymptotic approximation.

The measurement of the signal strength can be converted into a measurement of the Higgs boson production cross-section times the $H \rightarrow \ell\ell\gamma$ branching ratio in the fiducial region $m_{\ell\ell} < 30$ GeV. In this case, the included theory uncertainties are adjusted as discussed in Section 7, and the acceptance is extracted with the simulated samples described in Section 3.

7. Systematic uncertainties

The total observed systematic uncertainty in the signal strength measurement is 11%, which is about 35% of the size of the statistical uncertainty. Therefore, systematic uncertainties do not play a dominant role in this analysis.

The dominant experimental systematic uncertainties are due to the estimated biases in the fitted signal events (spurious signal, see Section 6). The corresponding uncertainty in the observed signal strength amounts to 6.1%. Other non-negligible systematic uncertainties relate to photon and lepton identification efficiencies in the simulated signal samples, in particular for merged-ee objects, as well as the energy/momentum calibration (see Section 4). Including jet uncertainties, which have a much smaller impact, these add up to 4.0%.

The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [95], obtained using the LUCID-2 detector [96] for the primary luminosity measurements. Uncertainties in the pile-up modelling contribute 1.7%. Uncertainties in the estimate of the $H \rightarrow \gamma\gamma$ branching fraction have a small impact of 0.7%. They include uncertainties in the photon–electron fake rate, and the uncertainties in the $\sigma(H) \times B(H \rightarrow \gamma\gamma)$ measurement [93].

The assumed uncertainty in the $H \rightarrow \gamma\gamma \rightarrow \ell\ell\gamma$ branching ratio contributes 5.8%. The choice of QCD scales impacts the number and distribution of signal events in the different categories and is evaluated for the ggF, VBF, and VH production modes using a scheme similar to the one discussed in Ref. [93]. The corresponding uncertainty in the measured signal strength is 4.7%. Uncertainties in the PDF are evaluated using the eigenvectors of the PDF4LHC15 PDF set [32] and have a smaller effect of 2.3%. A conservative uncertainty of 50% is assigned to the normalisation of the $t\bar{t}H$, $gg\rightarrow 2H$, and $bHb$ production modes, as no dedicated $H \rightarrow \gamma\gamma \rightarrow \ell\ell\gamma$ samples are available, with an impact of 0.8%. Parton shower uncertainties contribute only 0.3%.

For the $\sigma(H) \times B(H \rightarrow \ell\ell\gamma)$ measurement, the effect of the theory uncertainties is reduced to 1.1% for the QCD scale and 0.9% for the PDF uncertainty, as only acceptance effects are considered, whereas the uncertainties in the predicted cross-sections and branching ratio are not applicable to this measurement.

8. Results

The $m_{\ell\ell}$ distributions of the selected events and the result of the global fit of the parametric signal-plus-background models to the data are shown in Fig. 3 for each event category.

The best-fit value of the signal-strength parameter is $\mu = 1.5 \pm 0.5 = 1.5 \pm 0.5 \text{ (stat.)} \times 0.5 \text{ (syst.)}$, while the corresponding expected SM value is $\mu_{\text{exp}} = 1.0 \pm 0.5 = 1.0 \pm 0.5 \text{ (stat.)} \times 0.7 \text{ (syst.)}$. The best-fit signal strength in the muon (electron) channel, obtained from a fit with two separate signal-strength parameters, is $\mu_{\mu\mu} = 1.9 \pm 0.7$ ($\mu_{ee} = 1.0 \pm 0.7$). Fig. 4 shows the results of the fit when the signal strength in each category is allowed to float independently. As anticipated in Table 1, the low-$p_T$ categories, especially in the $\mu\mu$ channel, have the smallest uncertainties. It can be seen that all categories yield results that are consistent with each other and with the result of the single-$\mu$ fit.

For illustration, Fig. 5 shows the $m_{\ell\ell}$ distribution, with every data event reweighted by a category-dependent weight $\ln(1 + S_{\text{data}}/B_{\text{data}})$, where $S_{\text{data}}$ is the number of signal events in the smallest window containing 90% of the expected signal and $B_{\text{data}}$ is the expected number of background events in the same window, which consists of the resonant $H \rightarrow \gamma\gamma$ background as well as the non-resonant background. The latter is estimated from fits to the data sidebands using the background models.

Fig. 3. $m_{\ell\ell\gamma}$ distributions of the selected events and the results of the global fit, for the VBF-enriched categories (a, b, c), the high-$p_T\tau$ categories (d, e, f), and the low-$p_T\tau$ categories (g, h, i). The ee-resolved categories are shown in the left column, the ee-merged categories in the middle and the $\mu\mu$ categories in the right column. The data are shown as the black circles with statistical uncertainties. The red curve shows the combined signal-plus-background model, the dashed black line shows the model of the non-resonant background component and the dotted blue line denotes the sum of the non-resonant background and the resonant $H \rightarrow \gamma\gamma$ background. The curves are obtained from the fit, i.e. they include the best-fit values of the parameter of interest and the nuisance parameters, including the spurious signal. The bottom panel shows the residuals of the data with respect to the non-resonant background component of the signal-plus-background fit.

The observed (expected) significance over the background-only hypothesis for a Higgs boson with a mass of 125.09 GeV is 3.2σ (2.1σ).

The Higgs boson production cross-section times the $H \rightarrow \ell\ell\gamma$ branching ratio for $m_{\ell\ell} < 30$ GeV is determined to be $8.7^{+2.8}_{-2.7}$ fb = $8.7 \pm 2.7$ (stat.)$^{+0.7}_{-0.6}$ (syst.) fb.
9. Conclusion

A search for the Higgs boson decaying into a low-mass pair of electrons or muons and a photon is presented. The analysis is performed using a data set recorded by the ATLAS experiment at the LHC with proton-proton collisions at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 139 fb$^{-1}$. For a Higgs boson with a mass of 125.09 GeV and $m_{\ell\ell} < 30$ GeV, evidence for the $H \rightarrow \ell\ell\gamma$ process is found with a significance of 3.2$\sigma$ over the background-only hypothesis, compared to an expected significance of 2.1$\sigma$. The best-fit value of the signal-strength parameter, defined as the ratio of the observed signal yield to the signal yield expected in the Standard Model, is $\mu = 1.5 \pm 0.5$. The Higgs boson production cross-section times the $H \rightarrow \ell\ell\gamma$ branching ratio for $m_{\ell\ell} < 30$ GeV is determined to be 8.7$^{+2.5}_{-1.6}$ fb. This result constitutes the first evidence for the decay of the Higgs boson into a pair of leptons and a photon, an important step towards probing Higgs boson couplings in this rare decay channel.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, and CEA/DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GRSF, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; JST and JSPS, Japan; CNRS, Morocco; NWO, Netherlands; RCN, Norway; MNISW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZS, Slovenia; DST/NRF, South Africa; MICINN, Spain; 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, CANARIE, Compute Canada, CRC and IVADO, Canada; Beijing Municipal Science & Technology Commission, China; COST, ERC, ERDF, Horizon 2020 and Marie Sklodowska-Curie Actions, European Union; Investissements d'Avenir Labex, Investissements d'Avenir Idex and ANR, France; DFG and AVH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by ELI-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafsson Stiftelser, Sweden; 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 (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [97].

References

Also at Department of Physics, University of Fribourg, Fribourg, Switzerland.

Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia.

Also at Faculty of Physics, Sofia University, 'St. Kliment Ohridski', Sofia, Bulgaria.

Also at Giresun University, Faculty of Engineering, Giresun, Turkey.

Also at Graduate School of Science, Osaka University, Osaka, Japan.

Also at Hellenic Open University, Patras, Greece.

Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain.

Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany.

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary.

Also at Institute of Particle Physics (IPP), Canada.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan.

Also at Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid, Spain.

Also at Istanbul University, Dept. of Physics, Istanbul, Turkey.

Also at Joint Institute for Nuclear Research, Dubna, Russia.

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia.

Also at National Research Nuclear University MEPhI, Moscow, Russia.

Also at Physics Department, An-Najah National University, Nablus, Palestine.

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.

Also at The City College of New York, New York NY, United States of America.

Also at TRIUMF, Vancouver BC, Canada.

Also at Universita di Napoli Parthenope, Napoli, Italy.

Also at University of Chinese Academy of Sciences (UCAS), Beijing, China.

* Deceased.