Measurement of the Higgs boson mass in the HZZ*4 decay channel using 139 fb⁻¹ of s=13 TeV pp collisions recorded by the ATLAS detector at the LHC

Aad, G.; Abbott, B.; Abbott, DC; Abeling, K.; Abidi, S.H.; Aboulhorma,, A; Abramowicz, H.; Abreu, H.; Abulaiti, Y.; Abusleme Hoffman, A. C.; Dam, Mogens; Camplani, Alessandra; Hansen, Jørn Dines; Hansen, Jørgen Beck; Hansen, Peter Henrik; Ignazzi, Rosanna; Petersen, Troels Christian; Wiglesworth, Graig; Xella, Stefania; ATLAS

Published in:
Physics Letters B

DOI:
10.1016/j.physletb.2023.137880

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel using 139 fb$^{-1}$ of $\sqrt{s} = 13$ TeV $pp$ collisions recorded by the ATLAS detector at the LHC

The ATLAS Collaboration

© 2023 The Author(s). 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 SCOAP3.

1. Introduction

The observation of a Higgs boson, $H$, in 2012 by the ATLAS and CMS collaborations [1,2] in proton–proton ($pp$) collisions produced by the Large Hadron Collider (LHC) was a major step towards understanding the mechanism of the electroweak (EW) symmetry breaking [3–5]. An unknown parameter of the Standard Model (SM), the Higgs boson mass $m_H$, is related to the SM vacuum stability [6] and its value is required for precise calculations of EW observables, including the production and decay properties of the Higgs boson itself. These calculations are needed to test the cou-
pling structure of the SM Higgs boson, as suggested in Ref. [7] and references therein. Therefore, it is important to determine the Higgs boson mass experimentally.

The mass of the Higgs boson was measured to be $125.09 \pm 0.24$ GeV [8] in a combined analysis performed by the ATLAS and CMS collaborations using approximately 25 fb$^{-1}$ of $\sqrt{s} = 7$ and 8 TeV $pp$ collision data recorded in 2011 and 2012, respectively, commonly referred to as Run 1. The individual measurements, reported in Refs. [9,10], used the $H \to ZZ^* \to 4\ell$ ($\ell = e$ or $\mu$) and $H \to \gamma\gamma$ decay modes because of their excellent mass resolution.

The ATLAS Collaboration reported a measurement of $m_H$ using the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with 36.1 fb$^{-1}$ of $\sqrt{s} = 13$ TeV $pp$ collision data recorded in 2015 and 2016. This was combined with the Run 1 ATLAS measurement to obtain a value of $m_H = 124.97 \pm 0.24$ GeV [11]. The CMS Collaboration measured $m_H$ using the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with 35.9 fb$^{-1}$ of $\sqrt{s} = 13$ TeV $pp$ collision data recorded in 2016. This was combined with the Run 1 CMS measurement to obtain a value of $m_H = 125.38 \pm 0.14$ GeV [12].

This paper reports a new measurement of $m_H$ in the $H \to ZZ^* \to 4\ell$ channel using a 139 fb$^{-1}$ dataset of 13 TeV $pp$ collisions produced by the LHC and recorded by the ATLAS detector between 2015 and 2018, commonly referred to as Run 2. Most of the techniques used in this analysis are described in detail in Ref. [13].

The $H \to ZZ^* \to 4\ell$ process is identified by selecting four leptons ($\ell = e, \mu$) in the final state, and by identifying one pair of same-flavour leptons that is consistent with arising from an on-shell Z-boson decay. The mass measurement is performed using an analytic model to describe the distribution of the reconstructed 4\ell invariant mass, $m_{4\ell}$, as a function of $m_H$ and to fit to the observed distribution. In comparison with the previous ATLAS result from the $H \to ZZ^* \to 4\ell$ channel [11], the measurement has been improved by using the complete Run 2 dataset and a new high-precision muon momentum calibration. The $m_H$ uncertainty has also been reduced by introducing a neural-network-based classifier to discriminate between signal and background processes and by including the event-by-event invariant-mass resolution of the four-lepton system in the analytic model used to fit the collision data.

This paper is organised as follows. The ATLAS detector is described in Section 2, and the data samples and Monte Carlo simulations are presented in Section 3. The reconstruction of muons and electrons, as well as their calibration procedures, is discussed in Section 4, while the event selection is explained in Section 5. The statistical models describing the signal and background samples are defined in Section 6 and results of the measurement are presented in Section 7. The conclusions are provided in Section 8.

2. ATLAS detector

The ATLAS experiment [14] at the LHC is a multi-purpose particle detector with nearly 4\pi coverage in solid angle. It includes an inner tracking detector (ID) used for charged-particle tracking, surrounded by a 2 T solenoid. Electromagnetic (EM) and hadronic calorimeters are placed outside the solenoid, followed by a muon spectrometer (MS). The ID provides precise reconstruction of tracks within a pseudorapidity range $|\eta| \leq 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [15,16]. It is followed by the silicon microstrip tracker, which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron-identification information based on the detection of transition radiation X-ray photons.

The ATLAS calorimeter system covers the pseudorapidity range $|\eta| < 4.9$, with finer granularity over the region matching the inner detector. The lead/liquid-argon (LAr) EM calorimeter is divided into two half-barrels ($|\eta| < 1.475$) and two endcap components ($1.375 < |\eta| < 3.2$). It is segmented into three longitudinal (depth) sections over the region $|\eta| < 2.5$, and into two depth sections for $2.5 < |\eta| < 3.2$. Hadronic calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters extending the coverage to $|\eta| = 3.2$. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements, respectively.

The muon spectrometer, located beyond the calorimeters, is designed to detect muons in the region $|\eta| < 2.7$ and to provide momentum measurements with a relative resolution better than 3% over a wide range of transverse momentum, $p_T$, and up to 10% at $p_T \sim 1$ TeV. A system of three superconducting air-core toroidal magnets provides a magnetic field with a bending power of approximately 2.5 Tm in the barrel and up to 6 Tm in the endcaps. A two-level trigger system [17] is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a maximum rate of about 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average. An extensive software suite [18] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3. Data and event simulation

This measurement uses data from $pp$ collisions with a centre-of-mass energy of 13 TeV collected between 2015 and 2018 using single-lepton, dilepton, and trilepton triggers [19,20] as detailed in Ref. [13]. The combined efficiency of these triggers is approximately 98%, 99%, 97%, and 99% for the 4\mu, 2\mu2e, 2e2\mu, and 4e final states, respectively, for the simulated $H \to ZZ^* \to 4\ell$ events passing the event selection described in Section 5 (assuming $m_H = 125$ GeV). After data-quality requirements are imposed, the integrated luminosity of the data sample is 139 fb$^{-1}$ [21].

Details of the simulated Monte Carlo (MC) events used in this analysis can be found in Refs. [13,22]. Higgs boson production via the gluon–gluon fusion (ggF) process was modelled at next-to-next-to-leading-order (NLO) accuracy in the strong coupling constant $\alpha_s$ using the POWHEG NNLOPS generator [23–31] with the PDF4LHC15NLO set of parton distribution functions (PDFs) [32]. Higgs bosons produced via vector-boson fusion (VBF), in association with a vector boson (VH) or in association with a top-quark pair (ttH), were simulated at NLO accuracy with the POWHEG Box generator [25–27], using the PDF4LHC15NLO PDF set. The loop-induced $gg \toZH$ process was simulated with the POWHEG Box generator at leading-order (LO) accuracy. Higgs boson production in association with a top quark (tH) or with a bottom-quark pair (bbH) was simulated at NLO accuracy using the MadGraph5_AMC@NLO generator [33,34] with the NNPDF3.0 PDF
set [35]. For all signal processes, the EVTGEN 1.2.0 generator [36] was used for the simulation of the bottom- and charm-hadron decays. Correspondingly, the PYTHIA 8 generator [37] was used for the $H \to ZZ^* \to 4\ell$ decay as well as for parton showering, hadronisation, and simulation of the underlying event. Higgs bosons produced via the ggF and VBF processes or in association with a vector boson were simulated for a range of $m_H$ values from 123 to 127 GeV. Higgs boson production in association with a top quark, a top-quark pair, or a bottom-quark pair were simulated assuming $m_H = 125$ GeV, as these processes make a negligible contribution. These samples are normalised to cross-sections obtained from the most recent predictions provided by the LHC Higgs Working Group [7].

The $ZZ^*$ continuum background was modelled separately for quark-initiated (qqZZ), gluon-initiated (ggZZ), and vector boson scattering (EW ZZ) production. The qqZZ process was modelled at NLO accuracy in $\alpha_s$ using the SHERPA 2.2.2 generator [38-41] with the NNPDF2.3LO PDF set. The ggZZ process with 0 or 1 jet in the final state was modelled with a LO calculation using the same SHERPA generator setup, with higher-order corrections calculated using massless quark loops [42-44] in the heavy top-quark approximation [45]. The resulting ggZZ cross-section was scaled by $1.7 \pm 1.0$ to account for the uncertainty in additional higher-order effects. Production of EW ZZ was modelled at LO accuracy with the SHERPA v2.2.2 with matching of the matrix element to the parton shower using the ME+PS@NLO prescription [46]. These samples were normalised to cross-sections obtained directly from the SHERPA simulation. Alternative qqZZ samples, produced with the POWHEG BOX v2 and MADGRAPH5_AMC@NLO generators, are used for studies of the systematic uncertainties. For the nominal and alternative $ZZ^*$ continuum samples, the PYTHIA 8 generator was used for parton showering, hadronisation, and simulation of the underlying event.

Backgrounds from WZ production and $t\bar{t}$ events were modelled using the POWHEG BOX v2 generator, while Z bosons produced in association with jets were simulated using the SHERPA 2.2.1 generator. Minor contributions from processes with three electroweak bosons, denoted by $VVV$, were modelled using SHERPA 2.2.2. Small backgrounds originating from top-quark production in association with one or more electroweak bosons or additional top quarks, such as ttZ, tWZ, tWWW, tWZW, tZγ, ttZ, ttZ, tt, ttt, and tZZ (denoted by tXX), were simulated using the MADGRAPH5_AMC@NLO generator. The PYTHIA 8 generator was used for parton showering, hadronisation, and simulation of the underlying event.

Generated events were processed through the ATLAS detector simulation [47] within the GEANT4 framework [48] and reconstructed in the same way as collision data. Additional pp interactions in the same or neighbouring bunch crossings, referred to as pile-up, are included in the simulation. The pile-up events were modelled using the PYTHIA 8 generator with the A3 set of tuned parameters [49] and the NNPDF2.3LO PDF set [50].

The normalisation of the non-resonant $ZZ^*$ background component is determined by the fit of a signal and background model to the observed $m_{4\ell}$ distributions. The normalisations of the backgrounds from hadrons or hadron-decay products misidentified as prompt leptons, and from $Z+\text{jets}$, $t\bar{t}$ and WZ processes (referred to as the reducible background), are determined using data-driven techniques, and are explained in detail in Ref. [13].

4. Muon and electron reconstruction

Muon candidates are reconstructed using a combination of different algorithms [51]. The reconstruction of muon candidates within $|\eta| < 2.5$ is primarily performed by a global fit of reconstructed tracks in the ID and the MS. In the central detector region ($|\eta| < 0.1$), where the MS has reduced geometrical coverage, muons are also identified by matching a reconstructed ID track to either an MS track segment ('segment-tagged muons') or a calorimetric energy deposit consistent with a minimum-ionising particle ('calorimeter-tagged muons'). Calorimeter-tagged muons are required to have $p_T > 15$ GeV. For both the segment-tagged and calorimeter-tagged muons, the muon momentum is measured from the ID track alone. In the forward MS region ($2.5 < |\eta| < 2.7$) outside the ID coverage, MS tracks with hits in three MS layers are accepted as 'stand-alone muons' and combined with track segments formed from hits in the silicon tracker, if they exist. Additionally, 'loose' muon-identification criteria [51] are applied to reject low-quality tracks that have missing hits in the MS or have poor agreement between the reconstructed MS and ID tracks. These requirements have an efficiency of at least 98% for muons with $p_T > 5$ GeV. Muons are required to be isolated by using both calorimeter-based and track-based isolation variables and applying the ‘PflowLoose’ criteria [51]. The efficiency of the isolation selection is about 85% for muons with $p_T > 5$ GeV and increases to >97% for muons with $p_T > 20$ GeV.

Although the detector is well aligned overall, there are residual local misalignments that effect the reconstructed muon-track sagitta [52,53] and introduce a small, charge-dependent, momentum bias. The bias is measured with an iterative procedure that minimises the observed width of the line-shape of $Z \to \mu^+\mu^-$ decays [54]. This correction, which is applied to the collision data, reduces the width of the dimuon invariant-mass distribution in $Z$-boson decays by approximately 1% [54]. A systematic uncertainty is assigned to this correction to cover residual differences between observed and simulated data. It is estimated to be, on average, 1% of the correction itself for muons from $Z \to \mu^+\mu^-$ decays.

Corrections are also applied to the reconstructed momentum resolution and momentum scale are parameterised as an expansion in powers of the muon $p_T$, with each coefficient measured as a function of $\eta$ and $\phi$. The corrections are extracted from large data samples of $J/\psi \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ decays, using techniques that reduce any residual biases and by directly calibrating the global re-fitted track [54]. Compared to the previously used corrections, this reduces the associated uncertainties by approximately a factor of four.

The momentum-scale corrections range from 0.1% to 0.3% for muons with $p_T$ between 5 and 100 GeV. These account for inaccuracies in the estimation of the energy loss in the traversed material, local magnetic field inaccuracies, and geometrical distortions. The corrections to the momentum resolution for muons with $p_T$ of 5–100 GeV are at the percent level. The main sources of systematic uncertainty for these corrections are the residual biases in the calibration method and the consistency of calibrations using $J/\psi \to \mu^+\mu^-$ or $Z \to \mu^+\mu^-$ decays separately. For muons from $Z \to \mu^+\mu^-$ decays, which have an average $p_T$ of approximately 45 GeV, the momentum scale is determined with a precision better than 0.01%. The resolution is known with a precision of 0.4% for muons with $|\eta| < 1$, while for muons in high-$|\eta|$ regions the precision is approximately 1%.

A reconstructed electron consists of a cluster of energy deposits in the calorimeter and a matched ID track [55]. Variable-size clusters are created dynamically from calorimeter-energy deposits, improving the invariant-mass resolution of the four-lepton system, especially when bremsstrahlung photons are present. Electron ID tracks are fitted using an optimised Gaussian-sum filter (GSF) [56] that accounts for non-linear effects arising from energy loss through bremsstrahlung. Quality criteria are used to improve the purity of the selected electron candidates. The quality of an electron candidate is evaluated by using a likelihood method that employs measurements from the tracking system and the calorimeter system, and quantities that combine both tracking and calorimeter
information [56]. The ‘loose’ likelihood criteria, together with track hit requirements, are applied to electron candidates. Electrons are required to be isolated using both the calorimeter-based and track-based isolation variables as discussed in Ref. [13].

The energy of electrons is estimated from the calorimeter energy clusters, using a combination of simulation-based and data-driven corrections [55,57]. A single simulation-based correction, which accounts for the energy lost in the material upstream of the calorimeter, the energy deposited in the cells neighbouring the cluster in $\eta$ and $\phi$, and the energy lost beyond the LAr calorimeter, is derived using multivariate regression algorithms. Data-driven corrections account for effects such as those associated with the material in front of the EM calorimeter and material between the presampler and the calorimeter, and the inter-calibration of the different calorimeter layers. The remaining energy-scale difference between data and simulation is parameterised by energy-independent linear corrections, defined in different regions of $\eta$. Similarly, deviations of the energy resolution in the simulation from that in the data are parameterised as an $\eta$-dependent additional term. The electron reconstruction efficiency is $>97\%$ for electron $p_T > 18$ GeV and the additional identification efficiency varies between 75 and 90% for the kinematic region relevant to this analysis.

The main sources of systematic uncertainties in the electron-energy scale include uncertainties in the method used to extract the energy scale correction, as well as uncertainties due to the extrapolation of the energy scale from $Z \rightarrow e^+e^-$ events to electrons with different energies. A detailed explanation of these uncertainties can be found in Ref. [55]. In the case of electrons with $E_T$ around 40 GeV, the total relative uncertainties range between 0.04% and 0.2% for most of the detector acceptance. For electrons with $E_T$ around 10 GeV the relative uncertainty ranges between 0.3% and 0.8%.

Systematic uncertainties in the calorimeter-energy resolution for electrons arise from uncertainties in the modelling of the sampling term and in the measurement of the constant term in $Z$-boson decays, in the amount of material in front of the calorimeter, and in the modelling of the contribution to the resolution due to fluctuations in the pile-up from additional $pp$ interactions in the same or neighbouring bunch crossings. The uncertainty in the energy resolution for electrons with transverse energy between 30 and 60 GeV varies between 5% and 10%.

5. Event selection

Events are required to contain at least four isolated leptons emerging from a common vertex and forming two pairs of oppositely charged same-flavour leptons. Electrons are required to be within the geometrical acceptance of the inner detector ($|\eta| < 2.47$) and to have $E_T > 7$ GeV, while muons must be within the geometrical acceptance of the muon spectrometer ($|\eta| < 2.7$) and have $p_T > 5$ GeV (except for calorimeter-tagged muons, as explained in Section 4). At most one calorimeter-tagged or stand-alone muon is allowed per Higgs boson candidate. The three higher-($p_T$ or $E_T$) leptons in each quadruplet are required to pass thresholds of 20, 15, and 10 GeV, respectively, and have $\Delta R(\ell, \ell') > 0.1$. These thresholds are chosen to maximise signal yield consistent with selecting events with high trigger efficiency. Contributions from misidentified leptons are reduced by requiring the lepton tracks to have low transverse-impact-parameter significances and to be compatible with originating from a common vertex. A detailed description of the event selection can be found in Ref. [13].

The lepton pair with an invariant mass closest to the $Z$-boson mass in each quadruplet is referred to as the leading dilepton, while the remaining pair is referred to as the subleading dilepton. The selected quadruplets are separated into four subchannels, according to the flavour of the leading and subleading pairs. In order of decreasing expected selection efficiency and resolution, they are $4\mu, 2e+2\mu, 2\mu e2\mu$, and $4e$. The $m_{4\ell}$ resolution is $1.5$ GeV for subchannels with a subleading muon pair ($4\mu$ and $2e2\mu$) and about $2.1$ GeV for subchannels with a subleading electron pair ($2\mu e2\mu$ and $4e$). Only one quadruplet is selected from each event, based on the mass of the leading dilepton, the final state, and, for events with additional leptons, the value of the LO matrix element, as described in Ref. [13].

Final-state radiation (FSR) photons are searched for in all events following the procedure described in Ref. [13]. FSR candidates are defined as collinear if their angular separation from the nearest lepton of the quadruplet satisfies $\Delta R < 0.15$, and non-collinear otherwise. Collinear FSR candidates are considered only for muons from the leading dilepton as the electron candidate reconstruction includes collinear FSR, while non-collinear FSR candidates are considered for both muons and electrons from either the leading or the subleading dilepton. Only one FSR candidate is included in the quadruplet as events with more than one FSR candidate are rare and have negligible effect on the four-lepton mass resolution. In cases where there are more than one FSR candidate preference is given to collinear FSR and to the candidate with the highest $p_T$. FSR photons are found in 4% of the events and their energy is included in the mass computation, improving the $m_{4\ell}$ resolution by about 1%.

Finally, the four-momenta of leptons in the leading pair are recomputed by performing a kinematic fit which constrains their invariant mass to the $Z$-boson mass, as discussed in Ref. [58]. The fit, which takes into account lepton and FSR kinematic information, their associated experimental uncertainties, and the $Z$-boson width, improves the $m_{4\ell}$ resolution by about $17\%$ [9]. The lepton four-momenta reconstruction uncertainties were derived from simulation and are corrected to match those observed in data, using a large sample of $Z$-boson decays.

In the mass range $115 < m_{4\ell} < 130$ GeV, 313 candidate events are observed. The yield is in agreement with the expectation of $321 \pm 14$ events, 65% of which are predicted to originate from signal processes, assuming $m_{H^\pm} = 125$ GeV. The $m_{4\ell}$ distributions are shown in Fig. 1 for each final state and in Fig. 2 for the inclusive final state for the mass range of $105–160$ GeV, which is used to determine the Higgs boson mass.

The dominant contribution to the background is non-resonant $ZZ^\ast$ production, accounting for approximately 89% of the total background yield. The reducible background contributes approximately 9%. The $VVV$ and $t\bar{t}X$ events are estimated to constitute approximately 2% of the total background. The residual combinatorial background, originating from events with additional prompt leptons, is found to be negligible [13].

5.1. Signal–background discriminant

To provide additional separation between the $H \rightarrow ZZ^\ast \rightarrow 4\ell$ signal and the $ZZ^\ast \rightarrow 4\ell$ background, a deep feed-forward neural network (NN) is employed. The NN is trained with Keras [59] using the TensorFlow [60] backend, following the method described in Ref. [13]. Signal events for training are taken from simulated samples with different masses of the Higgs boson, as described in Section 3, thus reducing the dependence of the discriminant on $m_{4\ell}$.

The $p_T$ and $\eta$ of the four-lepton system are used as inputs to the NN, together with a matrix-element-based kinematic discriminant $D_{ZZ^\ast}$ [58]. The discriminant $D_{ZZ^\ast}$ is defined as $\ln[(\mathcal{M}_{HZZ^\ast})^2/\mathcal{M}_{ZZ^\ast}^2]$ where $\mathcal{M}_{HZZ^\ast}$ denotes the matrix element for leading-order (LO) $H \rightarrow ZZ^\ast \rightarrow 4\ell$ production and $\mathcal{M}_{ZZ^\ast}$ the sum of the corresponding matrix elements for the $q\bar{q} \rightarrow ZZ^\ast$
Fig. 1. The observed and expected (pre-fit) $m_{4\ell}$ distributions for the selected Higgs boson candidates, for the different decay final states (a) $4\mu$, (b) $2e2\mu$, (c) $2\mu2e$, and (d) $4e$ events. The predicted number of events for these distributions is taken from simulation for the signal, $ZZ^*$, $t\bar{t}X$, and VVV processes, while it is taken from the data-driven estimate (see Section 6) for the $Z$-jets and $t\bar{t}$ backgrounds. The total uncertainty in the prediction is shown by the hatched band, which also includes the theoretical uncertainties of the SM cross-section for the signal and the $ZZ^*$ background. Higgs boson events in this plot are simulated with $m_H = 125$ GeV. The $p$-values quantifying the agreement between the data and MC predictions are (a) 0.81, (b) 0.43, (c) 0.77, and (d) 0.15.

and $gg \to ZZ^*$ continuum backgrounds, all calculated at LO with the MadGraph 2.2.1 generator [33].

The distribution of the output of the NN, $D_{NN}$, is shown in Fig. 3. The $D_{NN}$ for each event is included in the final fit as an additional observable, as explained in Section 6. The additional separation between signal and background provided by the NN improves the precision of this measurement by 2%.

5.2. Event-level $m_{4\ell}$ resolution

The event-level $m_{4\ell}$ resolution, $\sigma_i$, is estimated using a quantile regression neural network (QRNN) [61]. The QRNN is trained on signal MC events using the Tensorflow library and the Keras Python application programming interface. The inputs are the $p_T$, $\eta$, and $\phi$ of the individual leptons, as well as the four-lepton momentum, constrained by the $Z$-boson mass constraint, and its uncertainty. The output of the QRNN for each final state is the predicted quantile of the difference between the reconstructed mass and the true mass of the four-lepton system. The targeted quantile is calibrated using a step-wise scanning procedure so that it results in an estimate that produces 68% confidence-level intervals when tested on simulated events.

The $\sigma_i$ distributions are shown in Fig. 4. Each event is assigned the $\sigma_i$ estimated by the QRNN and employed in the likelihood fit to $m_{4\ell}$. Taking into account the per-event resolution reduces the total expected uncertainty in $m_{4\ell}$ by 1%.

6. Signal and background model

The Higgs boson $m_{4\ell}$ distribution is the result of the convolution of its theoretical line-shape, a narrow relativistic Breit–Wigner (BW) distribution of 4.1 MeV width [7] centred on $m_H$, with the detector response for the four-lepton invariant-mass distribution. The BW width is more than two orders of magnitude smaller than the $m_{4\ell}$ detector resolution. Therefore, the signal line-shape in $m_{4\ell}$ is completely dominated by the detector response.
The ATLAS Collaboration
Physics Letters B 843 (2023) 137880

Fig. 2. The observed and expected (pre-fit) $m_{4\ell}$ distributions for the selected Higgs boson candidates. The predicted number of events for these distributions is taken from simulation for the signal, $ZZ^*$, $tX$, and $VVV$ processes, while it is taken from the data-driven estimate (see Section 6) for the $Z+\text{jets}$ and $t\bar{t}$ backgrounds. The total uncertainty in the prediction is shown by the hatched band, which also includes the theoretical uncertainties of the SM cross-section for the signal and the $ZZ^*$ background. Higgs boson events in this plot are simulated with $m_H = 125$ GeV. The $p$-value quantifying the agreement between the data and MC predictions is 0.28.

Fig. 3. The observed and expected (pre-fit) $D_{NN}$ distributions for the selected Higgs boson candidates. The predicted number of events for these distributions is taken from simulation for the signal, $ZZ^*$, $tX$, and $VVV$ processes, while it is taken from the data-driven estimate (see Section 6) for the $Z+\text{jets}$ and $t\bar{t}$ backgrounds. The total uncertainty in the prediction is shown by the hatched band, which also includes the theoretical uncertainties of the SM cross-section for the signal and the $ZZ^*$ background. Higgs boson events in this plot are simulated with $m_H = 125$ GeV. The $p$-value quantifying the agreement between the data and MC predictions is 0.28.

The signal probability density function is modelled as a function of $m_{4\ell}$, signal versus $ZZ^*$ discriminator $D_{NN}$, and the per-event resolution estimator $\sigma_i$. The model can be written as

$$P(m_{4\ell}, D_{NN}, \sigma_i|m_H) = P(m_{4\ell}|D_{NN}, \sigma_i, m_H) \cdot P(D_{NN}|\sigma_i, m_H) \cdot P(\sigma_i|m_H) \simeq P(m_{4\ell}|D_{NN}, \sigma_i, m_H) \cdot P(D_{NN}|m_H),$$

where the following approximations are used:

- $P(D_{NN}|m_H) \simeq P(D_{NN}|m_H)$ because the neural network discriminator does not directly depend on the per-event $m_{4\ell}$ resolution,
- $P(\sigma_i|m_H) \simeq P(\sigma_i)$ since the averaged per-event resolution does not depend on $m_H$ within the range of 105 to 160 GeV used in this measurement, and
- $P(\sigma_i)$ is omitted from the probability density function because it was observed that it has approximately the same distribution for signal and background events in the Higgs boson peak region. Dedicated checks of the assumption that $P(\sigma_i)$ can be omitted have shown that this has negligible impact on the measurement.

The probability density function $P(D_{NN}|m_H)$ is determined through interpolation between neighbouring $m_H$ values using simulated data. The probability density function $P(m_{4\ell}|D_{NN}, \sigma_i, m_H)$ in each subchannel is described by a double-sided Crystal Ball [62] probability density function that consists of a Gaussian core and two power-law tails.

The mean of the Gaussian core is parameterised as a function of $m_H$ and $D_{NN}$ for each subchannel $\lambda$ as

$$a^\lambda \cdot (m_H - 125 \text{ GeV}) + b^\lambda(D_{NN}),$$

in order to decorrelate the uncertainties of the shifts in the mean reconstructed $m_{4\ell}$ due to the subchannel dependence from those due to the $D_{NN}$ dependence. The $a^\lambda$ parameters are consistent with unity within a few percent for all final states. The dependence of $b^\lambda$ on $D_{NN}$ is parameterised as a second-order polynomial, and its values lie in the 124–125 GeV range for the $2\mu2e$ and $4e$ final states, and between 124.6 GeV and 125 GeV for the $2e2\mu$ and $4\mu$ final states. The residual correlations between the $m_H$ dependence of the mean and shape of the reconstructed $m_{4\ell}$ distributions, and the $D_{NN}$ dependence of the mean have been checked with simulation and found to be negligible. The $a^\lambda$ parameters and the $b^\lambda$ functions are extracted from an unbinned maximum-likelihood fit to the combined ggF, VBF, and $VH$ samples, which were simulated with different values of $m_H$, as described in Section 3.
The standard deviation of the Gaussian core is expressed as a function of the predicted event-level resolution $\sigma_i$ and is parameterised as a function of $D_{NN}$ to account for a residual correlation between $D_{NN}$ and $\sigma_i$. The parameters of the left tail of the double-sided Crystal Ball are parameterised as a function of $D_{NN}$ only, which results in few percent variation as a function of $D_{NN}$, while those of the right tail are found to be constant for a given final state.

The parameters of the signal model for each subchannel are derived by a simultaneous maximum-likelihood fit to signal MC samples for $m_H$ values ranging from 123 GeV to 127 GeV. The resulting probability density function is shown in Fig. 5 for the $m_H = 125$ GeV signal sample. The residual bias in the measured $m_H$ for the derived model, estimated by pseudo-experiments sampling MC events, is at the level of 6 MeV and is treated as an additional systematic uncertainty. In the fit of the model to the data, $m_H$ can vary freely while the other parameters are constrained to vary in accordance with the impact of the relevant systematic uncertainties. Negligible biases in $m_H$ are observed when fitting the model to simulated Higgs signal samples with different values of $m_H$.

The probability density functions for the $ZZ^*$ and $tX + VVV$ backgrounds are estimated as a function of $m_{4l}$ and $D_{NN}$ for each subchannel from MC simulation. The size of the background contributions from mismeasured and non-isolated leptons is estimated from a data-control region [13], with the dependence on $m_{4l}$ and $D_{NN}$ parameterised using simulated samples.

The probability density functions for the backgrounds are smoothed using an adaptive kernel estimation technique, employing Gaussian kernels [63] to reduce statistical fluctuations. The effects of uncertainties that modify the shapes of probability density functions are described by using non-linear moment morphing [64] to interpolate between the nominal probability density functions and those accounting for the relevant systematic variations. Such uncertainties include the theoretical and experimental uncertainties for the $ZZ^*$ and $tX + VVV$ probability density functions and imperfect knowledge of the relative background composition for the reducible background’s probability density function.

The $m_{4l}$ distribution is described by the sum of a signal distribution and the distributions of the backgrounds. The normalisations of the signal contribution and the $ZZ^*$ process in each subchannel are free parameters and extracted directly from the fit to data, while those of the reducible background in each channel are constrained according to estimates obtained from data, using minimal input from simulation and following the methodology described in Ref. [13]. The normalisation of the $tX + VVV$ background is constrained according to the uncertainties in the theory prediction.

The fit model was validated by utilising a two-step unblinding process, where the same, unknown, bias is applied to all events in the observed dataset and the likelihood model, as described in Section 7, is fitted to it. No significant changes to the compatibility between channels or to any free or constrained parameters, with the exception of $m_H$, are observed when this bias is removed.

### 7. Results

The mass measurement is performed by maximising the profile-likelihood ratio [65,66]

$$
\lambda(m_H) = -2 \ln \left( \frac{L(m_H, \hat{\theta}(m_H))}{L(\hat{m}_H, \hat{\theta})} \right),
$$

where $\hat{m}_H$ and $\hat{\theta}$ denote the unconditional maximum-likelihood estimates of the parameters of the likelihood function $L$, while $\hat{\theta}$ is the conditional maximum-likelihood estimate of the parameters $\theta$ for a fixed value of the parameter of interest, $m_H$. Systematic uncertainties and their correlations are modelled by introducing nuisance parameters $\theta$ with priors described by Gaussian or log-normal functions that reflect the uncertainties in the values of the nuisance parameters.

The estimate of $m_H$ is extracted by performing a simultaneous unbinned maximum-likelihood fit to the four subchannels ($4\mu$, $2e2\mu$, $2\mu2e$, and $4e$) in the $m_{4l}$ range between 105 and 160 GeV. The expected and observed yields, along with the signal-to-background ratio, are shown in Table 1 for the signal region defined by limiting the data to $m_{4l}$ between 115 and 130 GeV.

The free parameters of the fit are $m_H$, the signal normalisation, and background normalisations for each of the four subchannels, while the nuisance parameters associated with the systematic uncertainties are constrained by their respective priors. Fig. 6 shows the $m_{4l}$ distribution of the data together with the result of the fit to the $H \to ZZ^* \to 4l$ candidates. The small shape fluctuations in the background prediction arise from the integrations of the background probability density functions; their effects on the mass fit are negligible. The fit results in

<table>
<thead>
<tr>
<th>Final state</th>
<th>Higgs</th>
<th>ZZ, tXX, VVV</th>
<th>Reducible backgrounds</th>
<th>Expected total yield</th>
<th>Observed yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>4$\mu$</td>
<td>78 ± 5</td>
<td>38.7 ± 2.2</td>
<td>2.84 ± 0.17</td>
<td>120 ± 5</td>
<td>115</td>
</tr>
<tr>
<td>2$e2\mu$</td>
<td>53.4 ± 3.2</td>
<td>26.7 ± 1.4</td>
<td>3.02 ± 0.19</td>
<td>81.1 ± 3.5</td>
<td>94</td>
</tr>
<tr>
<td>2$\mu2e$</td>
<td>41.2 ± 3.0</td>
<td>15.3 ± 1.3</td>
<td>3.4 ± 0.5</td>
<td>62.5 ± 3.3</td>
<td>59</td>
</tr>
<tr>
<td>4e</td>
<td>36.2 ± 2.7</td>
<td>15.7 ± 1.6</td>
<td>2.83 ± 0.35</td>
<td>54.8 ± 3.2</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>209 ± 13</td>
<td>99 ± 6</td>
<td>12.2 ± 0.9</td>
<td>321 ± 14</td>
<td>313</td>
</tr>
</tbody>
</table>
\( m_{H} = 124.99 \pm 0.18 \text{(stat.)} \pm 0.04 \text{(syst.)} \) GeV.

All fitted normalisations are found to be compatible with the SM expectations. By allowing the fit to determine the normalisation in each subchannel, the result is insensitive to the SM predictions for each subchannel. There is negligible change in the fit results if the relative fractions are constrained to the SM. The fit for \( m_{H} \) is also performed independently for each decay channel. The \( m_{4\ell} \) distributions for the four channels are shown in Fig. 7. The resulting profile-likelihood-ratio variation as a function of \( m_{H} \) and the resulting fitted values are compared with those of the combined fit in Fig. 8. The \( m_{H} \) measurements from the individual channels are compatible with the combined measurement, with a \( p \)-value of 0.82.

The statistical uncertainty of \( m_{H} \) is estimated by fixing all nuisance parameters associated with systematic uncertainties to their best-fit values, leaving all remaining parameters unconstrained. Uncertainties on the Higgs boson signal originating from the QCD scale, PDF variations, and modelling of FSR account result in an uncertainty in \( m_{H} \) of 14 MeV. Other sources, including background modelling and simulation statistics, contribute less than 5 MeV. The systematic uncertainties are applied following the strategy detailed in Ref. [13]. The total systematic uncertainty is estimated by subtracting the square of the statistical uncertainty from the square of the total uncertainty and calculating the square root. Table 2 shows the leading contributions to the systematic uncertainty of \( m_{H} \). The uncertainties in the electron energy scale and in the muon momentum scale, resolution, and sagitta bias correction are described in Section 4. Other sources, including background modelling and simulation statistics, contribute less than 5 MeV. The total uncertainty is in agreement with the expectation of 0.19 GeV and is dominated by the statistical component. Fig. 9 shows the distribution of the expected uncertainty in \( m_{H} \) obtained by fitting a set of pseudo-experiments assuming \( m_{H} = 125 \) GeV.

The result of this measurement is combined with the measurement of \( m_{H} \) using the Run 1 dataset in the same final states, which was \( m_{H} = 124.51 \pm 0.52 \) [9]. The correlation scheme for the systematic uncertainties of the two measurements follows the one used in Ref. [11], where only the uncertainties in the lepton calibration were considered correlated. In this combination, the electron calibration uncertainty is correlated between the two measurements, while the muon calibration systematic uncertainty is uncorrelated due to the use of improved and independent techniques, as described in Section 4. Systematic uncertainties are reduced by about 14% when using this correlation scheme. The combined result, calculated by combining the likelihoods of the two measurements is

\[ m_{H} = 124.94 \pm 0.17 \text{(stat.)} \pm 0.03 \text{(syst.)} \) GeV,

which has slightly smaller statistical and systematic uncertainties. The \( p \)-value for the two individual results, assuming the Higgs boson mass is the combined value, is 0.4.

8. Summary

The mass of the Higgs boson is measured from a maximum-likelihood fit to the invariant mass and the predicted invariant-mass resolution of the \( H \rightarrow ZZ^{*} \rightarrow 4\ell \) decay channel. The results are obtained from the full Run 2 pp collision data sample recorded by the ATLAS experiment at the CERN Large Hadron Collider at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 139 fb\(^{-1}\). The measurement is based on the latest calibrations of muons and electrons, and on improvements to the analysis techniques used to obtain the previous result using data collected by ATLAS in 2015 and 2016.

The measured value of the Higgs boson mass for the \( H \rightarrow ZZ^{*} \rightarrow 4\ell \) channel is

\[ m_{H} = 124.99 \pm 0.18 \text{(stat.)} \pm 0.04 \text{(syst.)} \) GeV.

Thanks to a larger dataset, improved experimental techniques, and more precise lepton calibration, the statistical uncertainty of the measurement has been reduced by a factor of two and the systematic uncertainty by about 20% relative to the previous Run 2 published result.

This measurement is combined with the previous one obtained in the same channel with ATLAS Run 1 data. The result of the combination is

\[ m_{H} = 124.94 \pm 0.17 \text{(stat.)} \pm 0.03 \text{(syst.)} \) GeV.

This is the most precise measurement of \( m_{H} \) in the \( H \rightarrow ZZ^{*} \rightarrow 4\ell \) channel by the ATLAS Collaboration.

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.

Data availability

The authors are unable or have chosen not to specify which data has been used.
Fig. 7. The $m_4\ell$ data distribution (black points) for (a) $4\mu$, (b) $2e2\mu$, (c) $2\mu2e$, and (d) $4e$ events is shown along with the signal-plus-background post-fit probability density function (red line). The p-values quantifying the agreement between the data and the fit model are (a) 0.92, (b) 0.10, (c) 0.83, and (d) 0.97.

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; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRT, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEIN, Poland; FCT, Portugal; MNE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Knut and Alice Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and MINerva, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; 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. [67].

Fig. 8. The test statistic, $-2\ln(\lambda)$, values as a function of $m_H$ are shown in (a) for the fit in each of the final states $4\mu$ (purple), $2\mu 2\mu$ (green), $2\mu e$ (orange), and $4e$ (blue), and for the combined fit (black), both with (solid lines) and without (dashed lines) systematic uncertainties. The horizontal dashed line indicates the location of the one-σ uncertainty. The fit results obtained in each final state are shown in (b) together with the combined result. The uncertainty bar on each point is the total statistical and systematic uncertainty; the brackets show the statistical uncertainty only. The vertical (red) line indicates the combined Run 2 result, and the grey band its total uncertainty.

Fig. 9. The distributions of the total $m_H$ uncertainty from pseudo-experiments assuming $m_H = 125$ GeV are shown, for when the fit does (black) and does not (blue) take into account systematic uncertainties. The solid lines correspond to the expected uncertainty distribution from pseudo-experiments, while the vertical dashed lines indicate the observed values of the uncertainties. The one-sided $p$-value for compatibility between the observed and expected total uncertainties is 0.28.

### References


![Graph](https://example.com/graph.png)  
**Fig. 8.** The test statistic, $-2\ln(\lambda)$, values as a function of $m_H$ are shown in (a) for the fit in each of the final states $4\mu$ (purple), $2\mu 2\mu$ (green), $2\mu e$ (orange), and $4e$ (blue), and for the combined fit (black), both with (solid lines) and without (dashed lines) systematic uncertainties. The horizontal dashed line indicates the location of the one-σ uncertainty. The fit results obtained in each final state are shown in (b) together with the combined result. The uncertainty bar on each point is the total statistical and systematic uncertainty; the brackets show the statistical uncertainty only. The vertical (red) line indicates the combined Run 2 result, and the grey band its total uncertainty.

![Graph](https://example.com/graph.png)  
**Fig. 9.** The distributions of the total $m_H$ uncertainty from pseudo-experiments assuming $m_H = 125$ GeV are shown, for when the fit does (black) and does not (blue) take into account systematic uncertainties. The solid lines correspond to the expected uncertainty distribution from pseudo-experiments, while the vertical dashed lines indicate the observed values of the uncertainties. The one-sided $p$-value for compatibility between the observed and expected total uncertainties is 0.28.
The ATLAS Collaboration

Also Affiliated with an institute covered by a cooperation agreement with CERN.

Also at An-Najah National University, Nablus; Palestine.

Also at Borough of Manhattan Community College, City University of New York, New York NY; United States of America.

Also at Bruno Kessler Foundation, Trento; Italy.

Also at Center for High Energy Physics, Peking University; China.

Also at Centro Studi e Ricerche Enrico Fermi; Italy.

Also at CERN, Geneva; Switzerland.

Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève; Switzerland.

Also at Department of Financial and Management Engineering, University of the Aegean, Chios; Greece.

Also at Department of Physics and Astronomy, Michigan State University, East Lansing MI; United States of America.

Also at Department of Physics and Astronomy, University of Louisville, Louisville, KY; United States of America.

Also at Department of Physics, Ben Gurion University of the Negev, Beer Sheva; Israel.

Also at Department of Physics, California State University, East Bay; United States of America.

Also at Department of Physics, California State University, Sacramento; United States of America.

Also at Department of Physics, King’s College London, London; United Kingdom.

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

Also at Department of Physics, University of Thessaly; Greece.

Also at Department of Physics, Westmont College, Santa Barbara; United States of America.

Also at Hellespont Open University, Patras; Greece.

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

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

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

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

Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia.

Also at Lawrence Livermore National Laboratory, Livermore; United States of America.

Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen; Germany.

Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing; China.

Also at TRIUMF, Vancouver BC; Canada.

Also at Università di Napoli Parthenope, Napoli; Italy.

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

Also at University of Colorado Boulder, Department of Physics, Colorado; United States of America.

Also at Washington College, Maryland; United States of America.

Also at Yeditepe University, Physics Department, Istanbul; Türkiye.

* Deceased.