Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state is carried out with up to 36.1 fb$^{-1}$ of LHC proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector in 2015 and 2016. Three benchmark signals are studied: a spin-2 graviton decaying into a Higgs boson pair, a scalar resonance decaying into a Higgs boson pair, and Standard Model non-resonant Higgs boson pair production. Two analyses are carried out, each implementing a particular technique for the event reconstruction that targets Higgs bosons reconstructed as pairs of jets or single boosted jets. The resonance mass range covered is 260–3000 GeV. The analyses are statistically combined and upper limits on the production cross section of Higgs boson pairs times branching ratio to $b\bar{b}b\bar{b}$ are set in each model. No significant excess is observed; the largest deviation of data over prediction is found at a mass of 280 GeV, corresponding to 2.3 standard deviations globally. The observed 95% confidence level upper limit on the non-resonant production is 13 times the Standard Model prediction.

KEYWORDS: Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1804.06174

https://doi.org/10.1007/JHEP01(2019)030
1 Introduction

The discovery of the Standard Model (SM) Higgs boson ($H$) [1, 2] at the Large Hadron Collider (LHC) motivates searches for new physics using the Higgs boson as a probe. In particular, many models predict cross sections for Higgs boson pair production that are significantly greater than the SM prediction. Resonant Higgs boson pair production is predicted by models such as the bulk Randall-Sundrum model [3, 4], which features spin-2
Kaluza-Klein gravitons, $G_{KK}$, that subsequently decay to pairs of Higgs bosons. Extensions of the Higgs sector, such as two-Higgs-doublet models [5, 6], propose the existence of a heavy spin-0 scalar that can decay into $H$ pairs. Enhanced non-resonant Higgs boson pair production is predicted by other models, for example those featuring light coloured scalars [7] or direct $t\bar{t}HH$ vertices [8, 9].

Previous searches for Higgs boson pair production have all yielded null results. In the $bb\bar{b}\bar{b}$ channel, ATLAS searched for both non-resonant and resonant production in the mass range 400–3000 GeV using 3.2 fb$^{-1}$ of 13 TeV data [10] collected during 2015. CMS searched for the production of resonances with masses 750–3000 GeV using 13 TeV data [11] and with masses 270–1100 GeV with 8 TeV data [12]. Using 8 TeV data, ATLAS has examined the $bb\bar{b}\bar{b}$ [13], $b\bar{b}\gamma\gamma$ [14], $bb\tau^{+}\tau^{-}$ and $W^{+}W^{-}\gamma\gamma$ channels, all of which were combined in ref. [15]. CMS has performed searches using 13 TeV data for the $bb\tau^{+}\tau^{-}$ [16] and $bb\ell\bar{\ell}\nu\nu$ [17] final states, and used 8 TeV data to search for $bb\gamma\gamma$ [18] in addition to a search in multilepton and multilepton+photons final states [19].

The analyses presented in this paper exploit the decay mode with the largest branching ratio, $H \rightarrow b\bar{b}$, to search for Higgs boson pair production in both resonant and non-resonant production. Two analyses, which are complementary in their acceptance, are presented, each employing a unique technique to reconstruct the Higgs bosons. The resolved analysis is used for $HH$ systems in which the Higgs bosons have Lorentz boosts low enough that four $b$-jets can be reconstructed. The boosted analysis is used for those $HH$ systems in which the Higgs bosons have higher Lorentz boosts, which prevents the Higgs boson decay products from being resolved in the detector as separate $b$-jets. Instead, each Higgs boson candidate consists of a single large-radius jet, and the presence of $b$-quarks is inferred using smaller-radius jets built from charged-particle tracks.

Both analyses were re-optimized with respect to the previous ATLAS publication [10]; an improved algorithm to pair $b$-jets to Higgs boson candidates is used in the resolved analysis, and in the boosted analysis an additional signal-enriched sample is utilized. The dataset comprises 2015 and 2016 data, corresponding to 27.5 fb$^{-1}$ for the resolved analysis and 36.1 fb$^{-1}$ for the boosted analysis, with the difference due to the trigger selections used. The results are obtained using the resolved analysis for a resonance mass between 260 and 1400 GeV, and the boosted analysis for masses between 800 GeV and 3000 GeV. The main background is multijet production, which is estimated from data; the sub-leading background is $t\bar{t}$, which is estimated using both data and simulations. The two analyses employ orthogonal selections, and a statistical combination is performed in the mass range where they overlap. The final discriminants are the four-jet and dijet mass distributions in the resolved and boosted analyses, respectively. Searches are performed for the following benchmark signals: a spin-2 graviton decaying into Higgs bosons, a scalar resonance decaying into a Higgs boson pair, and SM non-resonant Higgs boson pair production.
2 ATLAS detector

The ATLAS experiment [20] at the LHC is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4π coverage in solid angle.\textsuperscript{1} It consists of an inner tracking detector (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic (EM) and hadronic calorimeters, and a muon spectrometer (MS). The ID covers the pseudorapidity range |\eta| < 2.5. It consists of silicon pixel, silicon microstrip, and straw-tube transition-radiation tracking detectors. An additional pixel detector layer [21], inserted at a mean radius of 3.3 cm, improves the identification of b-jets [22]. Lead/liquid-argon (LAr) sampling calorimeters provide EM energy measurements. A steel/scintillator-tile hadronic 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. The muon spectrometer surrounds the calorimeters and includes three large superconducting air-core toroids. The field integral of the toroids ranges between 2.0 and 6.0 T m for most of the detector. The MS includes a system of precision tracking chambers and triggering chambers. A dedicated trigger system is used to select events [23]. The first-level trigger is implemented in hardware and uses the calorimeter and muon detectors to reduce the accepted event rate to 100 kHz. This is followed by a software-based high-level trigger that reduces the accepted event rate to 1 kHz on average.

3 Data and simulation

3.1 Data

This analysis is performed on two LHC pp collision datasets at \( \sqrt{s} = 13 \) TeV. Data were collected during stable beam conditions and when all relevant detector systems were functional. The integrated luminosity of the dataset collected during 2015 was measured to be 3.2 fb\textsuperscript{-1}. The second dataset was collected during 2016 and corresponds to an integrated luminosity of 24.3 fb\textsuperscript{-1} for the resolved analysis and 32.9 fb\textsuperscript{-1} for the boosted analysis.

The difference in integrated luminosity between the two analyses results from the choices of triggers. In the resolved analysis, a combination of b-jet triggers is used. Events were required to feature either one b-tagged jet [24, 25] with transverse momentum \( p_T > 225 \) GeV, or two b-tagged jets, either both satisfying \( p_T > 35 \) GeV or both satisfying \( p_T > 55 \) GeV, with different requirements on the b-tagging. Some triggers required additional non-b-tagged jets. Due to a change in the online b-tagging algorithm between 2015 and 2016, the two datasets are treated independently until they are combined in the final statistical analysis. After the selection described in section 5, this combination of triggers is estimated to be 65% efficient for simulated signals with a Higgs boson pair invariant mass.

---

\textsuperscript{1} ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates \( (r, \phi) \) are used in the transverse plane, \( \phi \) being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle \( \theta \) as \( \eta = -\ln \tan(\theta/2) \). Angular distance is measured in units of \( \Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} \).
$m_{HH}$, of 280 GeV, rising to 100% efficiency for resonance masses greater than 600 GeV. During 2016 data-taking, a fraction of the data was affected by an inefficiency in the online vertex reconstruction, which reduced the efficiency of the algorithms used to identify $b$-jets; those events were not retained for further analysis. This reduces the integrated luminosity of the 2016 dataset for the resolved analysis to 24.3 fb$^{-1}$. In the boosted analysis, events were selected from the 2015 dataset using a trigger that required a single anti-$k_t$ jet [26] with radius parameter $R = 1.0$ and with $p_T > 360$ GeV. In 2016, a similar trigger was used but with a higher threshold of $p_T > 420$ GeV. The efficiency of these triggers is 100% for simulated signals passing the jet requirements described in section 6, so the 2015 and 2016 datasets were combined into one dataset.

### 3.2 Signal models and simulation

Simulated Monte Carlo (MC) event samples are used in this analysis to model signal production and the background from $t\bar{t}$. The dominant multi-jet background is modelled using data-driven techniques, as described in sections 5.2 and 6.2.

$gg \rightarrow$ Scalar $\rightarrow HH \rightarrow b\bar{b}b\bar{b}$ events were generated at LO in QCD with MG5\_aMC\@NLO 2.2.3 interfaced with HERWIG++ [27] for parton-showering, hadronization and simulation of the underlying event. CT10 [28] PDF sets were used for MG5\_aMC\@NLO and CTEQ6L1 [29] for HERWIG++. The UE-EE-5-CTEQ6L1 set of tuned underlying-event parameters [30] was used. No specific model was considered for computing the scalar signal cross sections.

Signal $G_{KK} \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ events were generated at leading order (LO) with MG5\_aMC\@NLO 2.2.2 [31] interfaced with PYTHIA 8.186 [32] for parton-showering, hadronization and underlying-event simulation. The NNPDF2.3 LO parton distribution function (PDF) set [33] was used for both MG5\_aMC\@NLO and PYTHIA. The A14 set of tuned underlying-event parameters [34] was used. These signal samples were generated with $k/M_{Pl} = 1$ or 2, where $k$ is the curvature of the warped extra dimension and $M_{Pl} = 2.4 \times 10^{18}$ GeV is the effective four-dimensional Planck scale.

For the evaluation of theoretical uncertainties in the signal modeling, samples were produced with variations of the factorization and renormalization scales, PDF sets (following the prescription from ref. [35]) and shower generator. For the latter, scalar (spin-2) samples were produced that are interfaced to PYTHIA 8.186 rather than HERWIG++ (and vice versa).

The decay widths of these three resonance models differ. The scalar signals were generated with a width of 1 GeV, allowing a study of generic narrow-width scalar signals. The widths of the graviton signals depend on the resonance mass and the value of $k/M_{Pl}$. Relative to the resonance mass, they range from 3% (6%) at low mass to 13% (25%) at the highest mass for $k/M_{Pl} = 1$ ($k/M_{Pl} = 2$). The graviton samples were normalized using cross sections from ref. [36].

Resonant signal samples for the scalar and $k/M_{Pl} = 1$ graviton models were produced with masses in 10 GeV steps between 260 and 300 GeV, in 100 GeV steps up to 1600 GeV, in 200 GeV steps up to 2000 GeV, and in 250 GeV steps up to 3000 GeV. Signal samples
for the $k/M_{Pl} = 2$ graviton model were produced with the same spacings but omitting the masses of 270 GeV, 290 GeV and 2750 GeV due to the larger generated width.

SM non-resonant production of Higgs boson pairs via the gluon-gluon fusion process was simulated at NLO with \texttt{MG5\_aMC@NLO} [37], using form factors for the top-quark loop from HPAIR [38, 39] to approximate finite top-quark mass effects. The simulated events were reweighted to reproduce the $m_{HH}$ spectrum obtained in refs. [40, 41], which calculated the process at NLO in QCD while fully accounting for the top-quark mass. The cross section times branching ratio to the $bb\bar{b}\bar{b}$ final state, evaluated at next-to-next-to-leading order (NNLO) with the summation of logarithms at next-to-next-leading-logarithm (NNLL) accuracy and including top-quark mass effects at NLO is $11.3^{+0.9}_{-1.0}\text{fb}$ [40]. The uncertainty includes the effects due to renormalization and factorization scales, PDF set, $\alpha_S$, and the $H \rightarrow b\bar{b}$ branching ratio. In all signal samples, the mass of the Higgs boson ($m_H$) was set to 125 GeV.

Interference effects between $HH$ resonant production and SM non-resonant $HH$ production are not included in the simulated samples.

The generation of $t\bar{t}$ events was performed with \textsc{Powheg-box v1} [42] using the CT10 PDF set. The parton shower, hadronization, and the underlying event were simulated using \textsc{Pythia} 6.428 [43] with the CTEQ6L1 PDF set and the corresponding Perugia 2012 set of tuned underlying-event parameters [44]. The top-quark mass was set to 172.5 GeV. Higher-order corrections to the $t\bar{t}$ cross section were computed with Top++ 2.0 [45]. These incorporate NNLO corrections in QCD, including resummation of NNLL soft gluon terms.

The $Z$+jets sample was generated using \textsc{Pythia} 8.186 with the NNPDF2.3 LO PDF set.

For all simulated samples, charm-hadron and bottom-hadron decays were handled by \textsc{EvtGen} 1.2.0 [46]. To simulate the impact of multiple $pp$ interactions that occur within the same or nearby bunch crossings (pile-up), minimum-bias events generated with \textsc{Pythia} 8 using the A2 set of tuned parameters [47] were overlaid on the hard-scatter event. The detector response was simulated with \textsc{Geant} 4 [48, 49] and the events were processed with the same reconstruction software as that used for the data.

4 Object reconstruction

Jets are built from topological clusters of energy deposits in calorimeter cells [50], using a four-momentum reconstruction scheme with massless clusters as input. The directions of jets are corrected to point back to the identified hard-scatter, proton-proton collision vertex, which is the vertex with the highest $\Sigma p_T^2$ of constituent tracks.

Jets are reconstructed using the anti-$k_T$ algorithm with different values of the radius parameter $R$. The jets with $R = 0.4$ (“small-$R$ jets”), used in the resolved analysis, are reconstructed from clusters calibrated at the electromagnetic (EM) scale [51]. The jets are corrected for additional energy deposited from pile-up interactions using an area-based correction [52]. They are then calibrated using $p_T$- and $\eta$-dependent calibration factors derived from simulation, before global sequential calibration [51] is applied, which reduces differences in calorimeter responses to gluon- or quark-initiated jets. The final calibration
is based on in situ measurements in collision data [53]. Jets with $p_T < 60$ GeV, $|\eta| < 2.4$, and with a large fraction of their energy arising from pile-up interactions are suppressed using tracking information, which was combined in a multivariate classification algorithm (jet vertex tagger) [54]. Events that pass a “medium” jet vertex tagger working point, corresponding to a 92% efficiency for jets at the EM scale with $20 < p_T < 60$ GeV, are retained in the analysis. Quality criteria are applied to the jets, and events with jets consistent with noise in the calorimeter or non-collision backgrounds are vetoed [55].

The jets with $R = 1.0$ (“large-$R$ jets”) used in the boosted analysis are built from locally calibrated [51] topological clusters. They are trimmed [56] to minimize the impact of energy deposits from pile-up interactions. Trimming proceeds by reclustering the jet with the $k_t$ algorithm [57] into $R = 0.2$ subjets and then removing those subjets with $p_{T\text{subjet}}/p_{T\text{jet}} < 0.05$, where $p_{T\text{subjet}}$ is the transverse momentum of the subjet and $p_{T\text{jet}}$ that of the original jet. The energy and mass scales of the trimmed jets are then calibrated using $p_T$- and $\eta$-dependent calibration factors derived from simulation [58]. The mass of the large-$R$ jets is computed using tracking and calorimeter information, also called the combined mass technique [59], which leads to a smaller mass resolution and better estimate of the median mass value than obtained using only calorimeter energy clusters.

Jets containing $b$-hadrons are identified using a score value computed from a multivariate $b$-tagging algorithm $MV2c10$ [24, 25], which makes use of observables provided by an impact parameter algorithm, an inclusive secondary vertex finding algorithm and a multi-vertex finding algorithm. The $MV2c10$ algorithm is applied to a set of charged-particle tracks that satisfy quality and impact parameter criteria and are matched to each jet. For large-$R$ jets, $b$-tagging is performed on anti-$k_t$ $R = 0.2$ track-jets [60] matched to the large-$R$ jets using ghost association [61]. These track-jets are required to be consistent with the primary vertex of the event as well as to satisfy $p_T > 10$ GeV and $|\eta| < 2.5$. The small radius parameter of the track-jets enables two nearby $b$-hadrons to be identified when their $\Delta R$ separation is less than 0.4, which is beneficial when reconstructing high-$p_T$ Higgs boson candidates. The $b$-tagging requirements of both the resolved and the boosted analyses use working points that lead to an efficiency of 70% for $b$-jets with $p_T > 20$ GeV when evaluated in a sample of simulated $t\bar{t}$ events. This working point corresponds to a rejection rate of jets originating from $u$, $d$- or $s$-quarks or gluons of 380 for the jets with $R = 0.4$ and 120 for the track-jets. The rejection of jets from $c$-quarks is 12 (7.1) for the $R = 0.4$ jets (track-jets).

Muons are reconstructed by combining tracks in the ID with tracks in the MS, and are required to have $p_T > 4$ GeV, $|\eta| < 2.5$ and to satisfy “medium” muon identification criteria [62]. If a muon is within $\Delta R = 0.4$ (0.2) of a jet used for $b$-tagging in the resolved (boosted) analysis, their four-momentum is added to the calorimeter-based jet’s four-momentum to partially account for the energy lost in semileptonic $b$-hadron decays.

5 Resolved analysis

The resolved analysis is optimized to discover signals that result in either non-resonant or low-mass resonant Higgs boson pair production. The strategy is to select two Higgs boson
candidates, each composed of two $b$-tagged anti-$k_t$ small-$R$ jets, with invariant masses near $m_H$.

The invariant mass of the two-Higgs-boson-candidate system ($m_{4j}$) is used as the final discriminant between Higgs boson pair production and the backgrounds (which are principally multijet, with some $t\bar{t}$). Resonant signals would lead to a localized excess, while non-resonant production would result in an excess in the tail of the $m_{4j}$ spectrum.

5.1 Selection

Events are required to have a primary vertex with at least two tracks matched to it. The selection proceeds with the requirement that the event contains at least four $b$-tagged anti-$k_t$ small-$R$ jets with $p_T > 40$ GeV and $|\eta| < 2.5$ (“four-tag” sample). The four jets with the highest $b$-tagging score are paired to construct two Higgs boson candidates. Initially, all possible pairings are considered. The angle between the decay products of the Higgs boson in the laboratory frame depends on the value of $m_{4j}$ and thus on the Lorentz boost of the Higgs boson. Accordingly, pairings of jets into Higgs boson candidates are only accepted if they satisfy the following requirements:

\[
\begin{align*}
\frac{360}{m_{4j}} - 0.5 &< \Delta R_{jj,\text{lead}} < \frac{653}{m_{4j}} + 0.475 & \text{if } m_{4j} < 1250 \text{ GeV}, \\
\frac{235}{m_{4j}} &< \Delta R_{jj,\text{subl}} < \frac{875}{m_{4j}} + 0.35 & \text{if } m_{4j} > 1250 \text{ GeV}.
\end{align*}
\]

In these expressions, $\Delta R_{jj,\text{lead}}$ is the angular distance between the jets in the leading Higgs boson candidate and $\Delta R_{jj,\text{subl}}$ for the sub-leading candidate. The leading Higgs boson candidate is defined as the candidate with the highest scalar sum of jet $p_T$. These requirements on $\Delta R_{jj}$ efficiently reject jet-pairings in which one of the $b$-tagged jets is not consistent with one originating from a Higgs boson decay. The specific numbers in this and the following selection requirements were chosen to maximize the sensitivity to the signal.

Events that have more than two Higgs boson candidates satisfying these requirements necessitate an algorithm in order to choose the correct pairs. In the absence of energy loss through semileptonic $B$-decays, the optimal choice would be the combination most consistent with the decays of two particles of equal mass.\(^2\)

To account for energy loss, the requirement of equal masses is modified. The distance, $D_{HH}$, of the pairing’s leading and sub-leading Higgs boson candidate’s masses, $(m_{2j}^{\text{lead}}, m_{2j}^{\text{subl}})$ from the line connecting (0 GeV, 0 GeV) and (120 GeV, 110 GeV) is computed, and the pairing with the smallest value of $D_{HH}$ is chosen. The values of 120 GeV and 110 GeV are chosen because they correspond to the median values of the narrowest intervals that contain 90% of the signal in simulation. The quantity $D_{HH}$ can be expressed

\(^2\)Explicitly requiring the masses to be equal to 125 GeV does not significantly increase signal efficiency, while it sculptures the background Higgs boson candidates’ mass distributions to look like those of the signal, reducing the signal vs background discrimination in these variables.
as follows:

\[ D_{HH} = \frac{|m_{2j}^{\text{lead}} - 120|}{\left| m_{2j}^{\text{subl}} \right|} \left/ \sqrt{1 + \left( \frac{120}{110} \right)^2} \right. \]

In signal simulation the pairing of jets (when four b-jets have been identified) is correct in at least 90% of the events, depending on \( m_{4j} \).

The resolved analysis searches for resonances with a wide range of masses, 260 GeV < \( m_{HH} < 1400 \) GeV, as well as non-resonant signals. Event selection criteria that vary as a function of \( m_{4j} \) are used to reject background and hence enhance the analysis sensitivity across this range. Mass-dependent requirements are imposed on the \( p_T \) of the leading Higgs boson candidate, \( p_T^{\text{lead}} \), and the sub-leading Higgs boson candidate, \( p_T^{\text{subl}} \):

\[ p_T^{\text{lead}} > 0.5 m_{4j} - 103 \text{ GeV}, \]
\[ p_T^{\text{subl}} > 0.33 m_{4j} - 73 \text{ GeV}. \]

A further (\( m_{4j} \)-independent) requirement is placed on the pseudorapidity difference between the two Higgs boson candidates, \( |\Delta \eta_{HH}| < 1.5 \), which rejects multijet events.

A requirement on the Higgs boson candidates’ masses is used to define the signal region:

\[ X_{HH} = \left( \frac{m_{2j}^{\text{lead}} - 120 \text{ GeV}}{0.1 m_{2j}^{\text{lead}}} \right)^2 + \left( \frac{m_{2j}^{\text{subl}} - 110 \text{ GeV}}{0.1 m_{2j}^{\text{subl}}} \right)^2 < 1.6, \tag{5.1} \]

where the 0.1\( m_{2j} \) terms represent the widths of the leading and sub-leading Higgs boson candidates’ mass distributions, derived from simulation. The signal region is shown as the inner region of figure 1.

To reduce the \( t\bar{t} \) background, hadronically decaying top-quark candidates are built from any three jets in the event, of which one must be a constituent of a Higgs boson candidate. These three jets are ordered by their b-tagging score. The highest one is considered as the b-jet originating from the top-quark candidate decay; the other two jets are considered as forming a hadronically decaying W boson candidate. A measure of the consistency of this combination with the top-quark hypothesis is then evaluated using the \( X_{Wt} \) variable:

\[ X_{Wt} = \left( \frac{m_W - 80 \text{ GeV}}{0.1 m_W} \right)^2 + \left( \frac{m_t - 173 \text{ GeV}}{0.1 m_t} \right)^2, \tag{5.2} \]

where \( m_W \) is the invariant mass of the two-jet W boson candidate and \( m_t \) that of the three-jet top candidate.

All possible combinations of three jets are considered and the top-quark candidate with the smallest \( X_{Wt} \) is chosen for each event. Events with the smallest \( X_{Wt} < 1.5 \) are vetoed in the final selection. This requirement reduces the \( t\bar{t} \) contamination where both top quarks decay without leptons (hadronic) by 60%, and the \( t\bar{t} \) events that contain leptons (semileptonic) by 45%.

A correction is made using the known Higgs boson mass, where each Higgs boson candidate’s four-momentum is multiplied by a correction factor \( m_H/m_{2j} \). This leads to an
improvement of approximately 30% in signal $m_{4j}$ resolution with a significant reduction of low-mass tails caused by energy loss and with little impact on the background $m_{4j}$ shape.

The fraction of signal events accepted by the detector multiplied by the efficiency of each selection step is shown in figure 2 for the narrow-width scalar, graviton, and SM non-resonant signal models. The acceptance times efficiency is higher for the graviton samples because spin-2 resonances decay more centrally, resulting in higher-$p_T$ jets. The acceptance times efficiency is limited at low mass by the $p_T$ requirement on the jets, and at high mass the chance to resolve four distinct jets becomes lower, and the $b$-tagging efficiency decreases.

5.2 Background estimation

After the full event selection is applied, about 95% of the background consists of multijet events, which are modelled using data. The remaining 5% are $t\bar{t}$ events. The $t\bar{t}$ background normalization is determined from data, while the $m_{4j}$ spectrum is taken from simulation. A data-driven estimate of $Z$+jets events yields a contribution of 0.2% to the total background, which is neglected in the following. Background from other sources, including processes involving single Higgs boson production, is also found to be negligible.

5.2.1 Multijet background

The multijet background is modelled with an independent data sample selected using the same trigger and selection requirements as used in the signal region, except for the $b$-tagging requirement: at least four jets with $p_T > 40$ GeV are required, and exactly two of them have to be $b$-tagged. This “two-tag” selection yields a data sample that consists of 88% multijet events and 12% $t\bar{t}$ events. The predicted signal contamination is negligible.
Figure 2. The selection acceptance times efficiency at each stage of the event selection for (a) a narrow-width scalar and (b) $G_{KK} \rightarrow HH \rightarrow bbb$ with $k/M_{Pl} = 1$ for a range of resonance masses and (c) the SM non-resonant signal. Each selection step is detailed in section 5.1.

To model the multijet background in the four-tag sample with events selected in the two-tag sample, a product of two weights is applied to the two-tag events, where each weight corrects for different effects: additional jet activity and kinematic differences from applying $b$-tagging requirements.

The weights are derived in a signal-depleted sideband region of the $m_{lead}^{2j} - m_{subl}^{2j}$ plane and validated using an orthogonal control region. The control region is defined as the region with

$$\sqrt{(m_{lead}^{2j} - 124 \text{ GeV})^2 + (m_{subl}^{2j} - 113 \text{ GeV})^2} < 30 \text{ GeV},$$

excluding the signal region defined in eq. (5.1). The sideband region is defined by

$$\sqrt{(m_{lead}^{2j} - 126 \text{ GeV})^2 + (m_{subl}^{2j} - 116 \text{ GeV})^2} < 45 \text{ GeV},$$

excluding the control and signal regions. The outer boundaries of the regions are selected to achieve sufficient statistical precision while ensuring that the kinematic properties of the sideband region are representative of the signal region. The shifted centres of these regions [cf. (120 GeV, 110 GeV) for the signal region] ensure the mean Higgs boson candidate masses are equal in the three regions.

The event weight that corrects for additional jet activity is obtained as follows. For each event, all possible combinations are considered where at least two anti-$b$-tagged jets, that pass the kinematics requirements, are treated as $b$-jets. A constant per-jet transfer factor, $f$, is assigned to each of these jets and a factor $(1-f)$ to the remaining ones, and a weight for the event is computed from the sum of all combinations. One of the combinations is then randomly chosen to form the Higgs boson candidates using the same procedure as in the four-tag sample. The transfer factor is determined by comparing the jet multiplicity distributions of the two-tag and four-tag selections in the sideband region. The resulting value is reported in table 1.
The events in the two-tag data sample are then weighted further to correct for the kinematic differences caused by the additional $b$-tagging requirement of the four-tag sample. These differences can arise for the following reasons: the $b$-tagging efficiency varies as a function of jet $p_T$ and $\eta$; the various multijet processes contribute with different fractions in each sample; and the fractions of events accepted by each trigger path changes.

The weights are determined by fitting cubic splines to the ratio of kinematic distributions of the total background model to data after subtracting the $t\bar{t}$ contribution in both samples (12% in the two-tag sideband and 7% in the four-tag sideband), before the cut on $X_{Wt}$. This is done for five distributions that show large differences between the four-tag and two-tag samples. These are: the average $|\eta|$ of the four jets constituting the Higgs boson candidate; the $p_T$ of the second and fourth leading (in $p_T$) constituent jets; the smallest $\Delta R$ between any two constituent jets; and the $\Delta R$ between the other two constituent jets. The reweighting is performed using one-dimensional distributions, and is iterated until the weights converged to stable values.

Agreement of the background model and data after these reweightings is checked in the control and sideband regions, and is found to be notably improved. This improvement is verified in the variables used to derive the weights and additionally in many other kinematic distributions.

5.2.2 Background normalization and the $t\bar{t}$ background

The $m_{4j}$ distribution of the $t\bar{t}$ background is modelled using simulation. To improve the statistical precision, the two-tag simulated distribution is used for the hadronic $t\bar{t}$ sample, after correction by the same kinematic weights used for the multijet model. This procedure is validated in an inclusive region that contains events from signal region, sideband and control region; and good agreement is observed between the corrected two-tag sample and the four-tag sample in several distributions. The semileptonic $t\bar{t}$ background is modelled directly, using the four-tag MC sample of the $t\bar{t}$ background.

The normalizations of the $t\bar{t}$ and multijet backgrounds are determined simultaneously by fitting the yields of semileptonic $t\bar{t}$, hadronic $t\bar{t}$, and multijet events in three background-enriched samples. These background-enriched samples are all defined as having Higgs boson candidate’s masses in the sideband region, but consistent results are found using the control region data. Specific background-enriched samples are defined with requirements additional to the sideband selection. The semileptonic enriched sample must contain an isolated muon [62] with $p_T > 25$ GeV and from eq. (5.2), $X_{Wt} < 1.5$. The sample enriched in hadronic $t\bar{t}$ requires $X_{Wt} < 0.75$, while the sample defined by $X_{Wt} > 0.75$ is enriched in multijet events.

There are three parameters in the normalization fit: $\mu_{\text{multijet}}$, which scales the multijet yield from the two-tag to the four-tag sideband region after the per-jet transfer factor $f$ and the kinematic weights have been applied; and two parameters $\alpha_{t\bar{t}}^{\text{hadronic}}$ and $\alpha_{t\bar{t}}^{\text{semileptonic}}$ that correct the normalizations of the yields for the hadronic and semileptonic $t\bar{t}$ MC samples in the four-tag sideband region.

The result of the normalization fit is presented in table 1. The uncertainties are those from limited data and MC sample sizes, and they are propagated to the final fit, as described in section 7.
Table 1. The fitted values of the normalization parameters $\mu_{\text{multijet}}$ and $\alpha$ of both $t\bar{t}$ samples and their statistical uncertainties, given for the two datasets. The per-jet transfer factor $f$ is also listed, and is explained in section 5.2.1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$f$</th>
<th>$\mu_{\text{multijet}}$</th>
<th>$\alpha_{\text{hadronic}}$</th>
<th>$\alpha_{\text{semileptonic}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>0.22</td>
<td>0.0838 ± 0.0038</td>
<td>1.19 ± 0.45</td>
<td>1.44 ± 0.48</td>
</tr>
<tr>
<td>2016</td>
<td>0.15</td>
<td>0.2007 ± 0.0031</td>
<td>1.15 ± 0.25</td>
<td>1.7 ± 0.19</td>
</tr>
</tbody>
</table>

Figure 3. Distributions of $m_{4j}$ in the control region of the resolved analysis for (a) 2015 data and (b) 2016 data, compared to the predicted backgrounds. The hatched bands represent the statistical uncertainties. The expected signal distributions of $G_{KK}$ resonances with masses of 800 and 1200 GeV, a 280 GeV scalar particle and SM non-resonant $HH$ production ($\times 100$) are also shown. The scalar sample is normalized to a cross section times branching ratio of 2.7 pb.

After the reweighting corrections and the application of the normalization, there is good agreement between the background model and the data distributions in the sideband as well as the control regions. The distributions of $m_{4j}$ in the control region for both datasets are displayed in figure 3.

5.3 Systematic uncertainties

Background uncertainties are propagated from the fit which determines the multijet and $t\bar{t}$ yields. The statistical uncertainties in the scale factors (table 1) are propagated including the correlations, by calculating three orthogonal eigenvariations from the covariance matrix of the normalization fit, resulting in three nuisance parameters, such that each parameter acts on the three background normalizations simultaneously.

Shape uncertainties in the multijet background are assessed by deriving an alternative background model using the same procedure as in the nominal case, but using data from the control region rather than from the sideband. This alternative model and the baseline are consistent with the observed data in their regions and with each other. The differences
between the baseline and the alternative are used as a background-model shape uncertainty, with a two-sided uncertainty defined by symmetrizing the difference about the baseline. The uncertainty is split into two components to allow two independent variations: a low-$H_T$ and a high-$H_T$ component, where $H_T$ is the scalar sum of the $p_T$ of the four jets constituting the Higgs boson candidates. The boundary value is 300 GeV. The low-$H_T$ shape uncertainty primarily affects the $m_{4j}$ spectrum below 400 GeV (close to the kinematic threshold) by up to 5%, and the high-$H_T$ uncertainty mainly $m_{4j}$ above this by up to 30% relative to nominal.

Shape uncertainties affecting the $t\bar{t}$ background component are dominated by those associated with the use of two-tag simulation to model the $m_{4j}$ distribution of hadronic $t\bar{t}$. Since this background is also reweighted, its uncertainties can be assessed using the same procedure as for the multijet background and are again split into low-$H_T$ and high-$H_T$ components. The impact of detector and theoretical modelling uncertainties on the $t\bar{t}$ background shape were assessed but are found to be negligible, because the data-driven reweighting procedure used for the multijet modelling compensates for biases in the $t\bar{t}$ sample by adjusting the multijet model.

Theoretical uncertainties in the signal acceptance result from variations of renormalization and factorization scales, PDF set uncertainties, and uncertainties in modelling of the underlying event and hadronic showers (therefore varying initial- and final-state radiation). The scales are varied by factors of 2 and 0.5. The PDF uncertainties are evaluated using PDF4LHC15 sets [35]. The parton shower and underlying event are varied by replacing HERWIG++ with PYTHIA for the scalar samples, and vice versa for the spin-2 samples. The total theoretical uncertainty is dominated by the shower variations. The size of the variation of the expected signal yield is typically below 10% but can increase to 23%, depending on the signal hypothesis.

The following detector modelling uncertainties are evaluated: uncertainties in the jet energy scale (JES) and resolution (JER), uncertainties in the $b$-tagging efficiency, and uncertainties in the trigger efficiency. The jet energy uncertainties are derived using in situ measurement techniques described in refs. [63–65]. The JES systematic uncertainty is evaluated following the prescription outlined in ref. [66]. The JER uncertainty is evaluated by smearing jet energies according to the systematic uncertainties of the resolution measurement [66].

The uncertainty in the integrated luminosity for the 2015 (2016) dataset is ±2.1% (±2.2%), which was evaluated using a technique similar to that described in ref. [67].

The uncertainty in the $b$-tagging efficiency is evaluated by propagating the systematic uncertainty in the $p_T$-dependent, measured tagging efficiency for $b$-jets [68]. For $b$-jets with $p_T > 300$ GeV, systematic uncertainties in the tagging efficiencies are extrapolated with simulation and are consequently larger [25]. Uncertainties arising from mis-tagging jets that do not contain $b$-hadrons are negligible.

Trigger efficiency uncertainties are evaluated for the signal, based on the systematic uncertainties arising from the per-jet online $b$-tagging measurements. There is an additional, small, non-closure uncertainty associated with the calculation of per-event trigger efficiencies using the measured per-jet efficiencies. The total trigger efficiency uncertainty
The number of events observed in the data, the predicted number of background events in the signal region, and the predicted yield for three potential signals are presented in table 3 for both the 2015 and 2016 datasets. The numbers of observed and predicted events in the control region are also given, and they are in agreement. A discrepancy between data and the total prediction is seen in the 2016 dataset; about half of this excess can be attributed to one bin at $m_{4j} = 280$ GeV.

Figure 4 shows comparisons of the predicted $m_{4j}$ background distributions to those observed in the 2015 and 2016 datasets. A few signal models are also displayed. The scalar sample shown is normalized to a cross section times $H \rightarrow bb$ branching ratio of 2.7 pb, which is the best-fit value (the fit is described in section 7). The predicted background and observed distributions are mostly in agreement.
Table 3. The number of predicted background events in the signal region (SR) for the resolved analysis compared to the data, for the 2015 and 2016 datasets. The yields for three potential signals, an 800 GeV $G_{KK}$ resonance with $k/M_{Pl} = 1$, a scalar with a mass of 280 GeV, and SM non-resonant Higgs boson pair production, are also shown. The scalar sample is normalized to a cross section times branching ratio of 2.7 pb. The quoted uncertainties include both the statistical and systematic uncertainties, and the total uncertainty considers correlations. The numbers of observed and predicted events are also given in the control region (CR).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multijet</td>
<td>866 ± 70</td>
<td>6750 ± 170</td>
<td>880 ± 71</td>
<td>7110 ± 180</td>
</tr>
<tr>
<td>$t\bar{t}$, hadronic</td>
<td>52 ± 35</td>
<td>259 ± 57</td>
<td>56 ± 37</td>
<td>276 ± 61</td>
</tr>
<tr>
<td>$t\bar{t}$, semileptonic</td>
<td>13.9 ± 6.5</td>
<td>123 ± 30</td>
<td>20 ± 9</td>
<td>168 ± 40</td>
</tr>
<tr>
<td>Total</td>
<td>930 ± 70</td>
<td>7130 ± 130</td>
<td>956 ± 50</td>
<td>7550 ± 130</td>
</tr>
<tr>
<td>Data</td>
<td>928</td>
<td>7430</td>
<td>969</td>
<td>7656</td>
</tr>
</tbody>
</table>

$G_{KK}$ (800 GeV) 12.5 ± 1.9 89 ± 14 
Scalar (280 GeV) 24.0 ± 7.5 180 ± 57 
SM $HH$ 0.607 ± 0.091 4.43 ± 0.66

Figure 4. Distributions of $m_{4j}$ in the signal region of the resolved analysis for (a) 2015 data and (b) 2016 data, compared to the predicted backgrounds. The hatched bands represent the combined statistical and systematic uncertainties in the total background estimates. The expected signal distributions of $G_{KK}$ resonances with masses of 800 and 1200 GeV, a 280 GeV scalar sample and SM non-resonant $HH$ production (×100) are also shown. The scalar sample is normalized to a cross section times branching ratio of 2.7 pb.
6 Boosted analysis

The boosted analysis is optimized to discover signals arising from production of high-mass resonances decaying into Higgs boson pairs. The strategy is to select two Higgs boson candidates with mass near $m_H$, each composed of a single large-$R$ jet with at least one $b$-tagged track-jet matched to it. Three samples are defined according to the total number of $b$-tagged track-jets associated with the Higgs boson candidates. Since the triggers are fully efficient for the signal processes in both the 2015 and the 2016 datasets, the two datasets are combined into one.

The invariant mass of the two-Higgs-boson-candidate system, $m_{2J}$, is used as the final discriminant between Higgs boson pair production and the SM backgrounds. Events that pass the resolved signal region selection are vetoed in the boosted analysis, thus priority is given to the resolved analysis if an event passes both selections, which increases the sensitivity.

6.1 Selection

Events are required to have a primary vertex with at least two tracks matched to it. Events are selected that have at least two anti-$k_t$ large-$R$ jets with $p_T > 250$ GeV, $|\eta| < 2.0$, and mass $m_J > 50$ GeV. Only the two jets with highest $p_T$ are retained for further selection. The leading jet is required to have $p_T > 450$ GeV, which ensures 100% trigger efficiency. Since high-mass resonances tend to produce jets that are more central than those from multijet background processes, the two large-$R$ jets are required to have a separation $|\Delta \eta| < 1.7$. To be considered as a Higgs boson candidate, each large-$R$ jet must contain at least one $b$-tagged $R = 0.2$ track-jet matched to it by ghost association.

Three separate samples of events are selected. The “two-tag” sample requires each Higgs boson candidate to have exactly one associated $b$-tagged track-jet. The “three-tag” and “four-tag” samples require that there are exactly three or exactly four $b$-tagged track-jets associated with Higgs boson candidates in the event, with two $b$-tagged track-jets associated with one candidate and one or two associated with the other candidate. Increasing the required number of associated $b$-tagged track-jets in the event increases signal purity at the expense of lower signal efficiency. This loss of efficiency is particularly pronounced for the highest resonance masses. It is rare to identify four distinct track-jets containing $b$-hadrons in these high-mass events due to the inefficiency in $b$-tagging high-$p_T$ track-jets, and also because the extremely high Lorentz boosts make it difficult to resolve each $b$-quark as a separate track-jet.

Signal event candidates are selected if each of the large-$R$ jets has a mass consistent with that of the Higgs boson. This is defined analogously to the resolved analysis in eq. (5.1):

$$X_{HH} = \sqrt{\left(\frac{m_{J\text{lead}} - 124 \text{ GeV}}{0.1 m_{J\text{lead}}}\right)^2 + \left(\frac{m_{J\text{subl}} - 115 \text{ GeV}}{0.1 m_{J\text{subl}}}\right)^2} < 1.6,$$

with the large-$R$ jet masses $m_{J\text{lead}}$ and $m_{J\text{subl}}$, where leading (sub-leading) refers to the large-$R$ jet with the largest (second largest) $p_T$. The mass resolution of the Higgs boson
Figure 5. (a) The selection acceptance times efficiency of the boosted analysis at each stage of the event selection as a function of the generated graviton mass for $k/M_{P1} = 1$. The trigger efficiency is approximately 100% after the requirement of two large-$R$ jets, so it is not shown. (b) The selection efficiency when requiring two, three or four $b$-tagged track jets associated to the two large-$R$ jets, as a function of the generated graviton mass for $k/M_{P1} = 1$. In both figures for the case of two $b$-tagged track jets, they can either both be associated to the same, or one to each large-$R$ jet.

candidates found in simulation is similar to that in the resolved analysis. The central values of 124 GeV and 115 GeV correspond to the median values of the narrowest intervals that contain 90% of the simulated signal. The requirement of $X_{HH} < 1.6$ is chosen such that it optimizes the sensitivity, and it defines the signal region in the leading-sub-leading Higgs boson candidate mass plane.

The fraction of signal events accepted by the detector multiplied by the efficiency of each selection step for the $G_{KK}$ model with $k/M_{P1} = 1$ is shown in figure 5a, and the efficiency for signal events to populate samples with either two, three or four $b$-tagged track jets is displayed in figure 5b.

A correction is made, multiplying the four-momentum of each large-$R$ jet by a factor $m_H/m_J$. This slightly improves the resolution of $m_{2J}$ for signal events by reducing the low-mass tails caused by energy loss. There is little impact on the background distribution.

6.2 Background estimation

The main backgrounds after selection are multijet events, which comprise 80–95% of the total background, depending on the number of $b$-tagged track-jets required, with the remainder being $t\bar{t}$ events. The contribution of $Z +$ jets events to the signal region in each sample is estimated using simulation to be less than 0.1%, and is therefore neglected. Other sources of background, including processes involving single Higgs boson production, are also negligible. The multijet events are modelled using data. The $t\bar{t}$ yield is estimated using a data-driven technique, while the $t\bar{t} m_{2J}$ shape is taken from simulation.

The shape of the multijet background is modelled using independent data samples selected with the same trigger and selection requirements as described above, but with fewer $b$-tagged track-jets. To model the two-tag sample, a “1b-1” sample is selected comprising
events where one of the large-$R$ jets contains a single $b$-tagged track-jet, while the other large-$R$ jet contains no $b$-tagged track-jets, but at least one track-jet which fails $b$-tagging. Analogous “2$b$-1” and “2$b$-2” samples are selected to model the three- and four-tag samples, respectively. These comprise events where one large-$R$ jet contains two $b$-tagged track-jets and the other large-$R$ jet contains no $b$-tagged track-jets but exactly one (2$b$-1) or at least two (2$b$-2) track-jets that fail $b$-tagging. These selections are referred to below as lower-tagged samples.

The normalizations of both the multijet and the $t\bar{t}$ backgrounds are derived using a signal-free sideband region. The definition of the sideband region is optimized such that it contains the bulk of the $t\bar{t}$ events, yet is close enough to the signal region to accurately model the background kinematics there. A control region between the signal and sideband regions is used to validate the background models and to assign systematic uncertainties. The regions are defined using $X_{HH}$ and two other variables:

$$R_{HH} = \sqrt{(m_{j}^{\text{lead}} - 124 \text{ GeV})^2 + (m_{j}^{\text{subl}} - 115 \text{ GeV})^2},$$

$$R_{HH}^{\text{high}} = \sqrt{(m_{j}^{\text{lead}} - 134 \text{ GeV})^2 + (m_{j}^{\text{subl}} - 125 \text{ GeV})^2}.$$

The signal region contains events with $X_{HH} < 1.6$, events in the control region fulfil $R_{HH} < 33$ GeV and $X_{HH} > 1.6$, and events in the sideband satisfy $R_{HH} > 33$ GeV and $R_{HH}^{\text{high}} < 58$ GeV. The three regions of the $m_{j}^{\text{lead}}$-$m_{j}^{\text{subl}}$ plane are depicted in figure 6.

Similarly to the resolved analysis, corrections are made for differences between the lower-tagged and $n$-tag samples by reweighting events in the lower-tagged sample. Differences between those samples are expected since requiring $b$-tags generally affects the jet
kinematics. In the 1b-1 sample the anti-\(b\)-tagged large-\(R\) jet’s kinematics are reweighted to mimic the kinematics of a tagged large-\(R\) jet (i.e. a Higgs boson candidate). Similarly, in the 2b-1 (2b-2) sample the anti-\(b\)-tagged large-\(R\) jet’s kinematics are reweighted to the kinematics of a Higgs boson candidate with one (two) \(b\)-tags. The weights are derived from data from lower-tagged samples which are inclusive in the regions, i.e. events from signal, control or sideband region. Each lower-tagged sample is split into two subsamples, depending on whether the leading or sub-leading large-\(R\) jet is \(b\)-tagged. The weights are obtained from spline interpolations to the ratios of the two subsamples for the three distributions that are most affected by \(b\)-tagging: the \(p_T\) of the leading large-\(R\) jet, and the \(p_T\) of the leading track-jets associated with the leading and sub-leading large-\(R\) jets. The reweighting is iterated until the weights converge to stable values.

The background yield in each of the three signal samples, \(N_{n\text{-tag}}\text{background}\) (where \(n\)-tag represents two-, three- and four-tag), is evaluated using the following expression:

\[
N_{n\text{-tag}}\text{background} = \mu_{n\text{-tag}}^{\text{multijet}} N_{n\text{-tag}}^{\text{lower-tag}} + \alpha_{n\text{-tag}}^{t\bar{t}} N_{n\text{-tag}}^{t\bar{t}} ,
\]  

(6.1)

where \(N_{n\text{-tag}}^{\text{multijet}}\) is the number of multijet events in the lower-tagged sample and \(N_{n\text{-tag}}^{t\bar{t}}\) are the numbers of events predicted by the \(n\)-tag \(t\bar{t}\) simulation. The parameter \(\mu_{n\text{-tag}}^{\text{multijet}}\) corresponds to the ratio of multijet event yields in the \(n\)-tag and lower-tagged samples. The parameter \(\alpha_{n\text{-tag}}^{t\bar{t}}\) is a scale factor designed to correct the \(t\bar{t}\) event yield estimated from the simulation in the \(n\)-tag samples.

The values of \(\mu_{n\text{-tag}}^{\text{multijet}}\) and \(\alpha_{n\text{-tag}}^{t\bar{t}}\) are extracted using eq. (6.1) from binned likelihood fits to the leading large-\(R\) jet’s mass distribution in the sideband region. The leading jet’s mass distributions (both normalization and shape) for multijet events are obtained from the lower-tagged samples, after subtraction of the \(t\bar{t}\) contributions predicted by simulation. The fitted values of \(\mu_{n\text{-tag}}^{\text{multijet}}\) and \(\alpha_{n\text{-tag}}^{t\bar{t}}\) are given in table 4.

The impact of limited statistical precision for \(m_{2J} > 1.2\) TeV in the multijet and \(t\bar{t}\) models is ameliorated by fitting the background distributions with the following function:

\[
f \left( m_{2J} \right) = \frac{p_1}{m_{2J}} \left( 1 - \frac{m_{2J}}{\sqrt{s}} \right)^{p_2-p_3 \ln \frac{m_{2J}}{\sqrt{s}}} ,
\]

(6.2)

where \(p_i\) are free parameters and \(\sqrt{s}\) is the centre-of-mass energy. This functional form is chosen after fitting various functions to the lower-tagged data. The \(m_{2J}\) distributions of the multijet and \(t\bar{t}\) backgrounds are each fitted in the range 1.2 TeV < \(m_{2J}\) < 3.0 TeV.

Table 4. The fitted values of \(\mu_{n\text{-tag}}^{\text{multijet}}\) and \(\alpha_{n\text{-tag}}^{t\bar{t}}\) for the two-tag, three-tag and four-tag samples. The statistical uncertainties are shown as well.

<table>
<thead>
<tr>
<th>Category</th>
<th>(\mu_{\text{multijet}})</th>
<th>(\alpha_{t\bar{t}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-tag</td>
<td>0.06273 ± 0.00057</td>
<td>0.986 ± 0.019</td>
</tr>
<tr>
<td>Three-tag</td>
<td>0.1626 ± 0.0043</td>
<td>0.800 ± 0.073</td>
</tr>
<tr>
<td>Four-tag</td>
<td>0.0332 ± 0.0043</td>
<td>0.89 ± 0.60</td>
</tr>
</tbody>
</table>
Table 5. The number of events in data and predicted background yields in the sideband and control regions of the two-tag, three-tag and four-tag samples for the boosted analysis. The numbers of multijet and $t\bar{t}$ background events in the sideband regions are constrained by the number of observed events, as explained in the text. The uncertainties are statistical, with fit uncertainties included for backgrounds. The anti-correlation between the multijet and $t\bar{t}$ yields is accounted for in the uncertainty in the total background yield.

From these fits, smooth background histograms are generated and passed to the statistical analysis. Since very few simulated $t\bar{t}$ events pass the full three-tag or four-tag selections, the shape of the $t\bar{t}$ distribution in the three-tag or four-tag sample is taken from the two-tag distribution. The shape differences of these templates are negligible compared to the systematic uncertainties considered and the statistical uncertainties of the available samples.

The modelling of the background yields and kinematics is validated in the control regions of the $n$-tag samples. Good agreement is observed between the yield in data and the yield of predicted backgrounds in the sideband and control regions of each of the samples, as shown in table 5. Figure 7 compares the predicted background $m_{2J}$ distributions to data in the control regions of the three samples. The systematic uncertainties derived in section 6.3 are shown.

6.3 Systematic uncertainties

Uncertainties that are in common with those of the resolved analysis are the theoretical uncertainties in the signal acceptance and the $b$-tagging uncertainties. Systematic uncertainties that differ from those of the resolved analysis are described here.

The uncertainty in the integrated luminosity of the combined 2015 and 2016 datasets is $\pm 2.1\%$ [67].

The large-$R$ jet energy and mass uncertainties (i.e. jet mass scale, JMS, and jet mass resolution, JMR) are derived in situ from 13 TeV $pp$ collisions, using techniques described in ref. [69]. The uncertainty in the $b$-tagging efficiency for track-jets is evaluated with the same method as used for $R = 0.4$ calorimeter-based jets.

Uncertainties in the signal are treated as fully correlated across the three samples. Detector modelling uncertainties in the $t\bar{t}$ sample ($b$-tagging efficiencies; jet energy, resolution and mass) impact the result of the normalization fit. These variations of $\mu_{\text{multijet}}$ and $\alpha_{\text{t\bar{t}}}$ are propagated to the predictions of the multijet and $t\bar{t}$ estimates in the signal regions, and they are treated as fully correlated across the three signal regions and are fully correlated with the same uncertainties in the signal.
An uncertainty in both the shape and normalization of the multijet and $t\bar{t}$ backgrounds is assigned by considering the statistical uncertainties in the nominal fitted values of $\alpha_{t\bar{t}}$ and $\mu_{\text{multijet}}$, as given in table 4. Two orthogonal eigenvariations are calculated from the covariance matrix of the normalization fit, which are then applied to the background predictions. Correlations between $\alpha_{t\bar{t}}$ and $\mu_{\text{multijet}}$ are fully retained this way.

An additional uncertainty in the shape of the multijet estimate, accounting for the choice of smoothing procedure, is obtained by comparing smoothed data to smoothed prediction in the control region, and assigning this difference as a shape uncertainty. The same function as defined in eq. (6.2) is used for these smoothings, which mitigates statistical fluctuations. The uncertainty is split at $m_{2J} = 2$ TeV into low- and high-$m_{2J}$ components.
Table 6. Summary of systematic uncertainties (expressed in percentage) in the total background and signal event yields in the signal region of the boosted analysis. Uncertainties are provided for each of the three samples for background, a 2 TeV scalar, and a $G_{KK}$ with $k = \sqrt{M_{Pl}} = 1$ and $m = 2.0$ TeV.

<table>
<thead>
<tr>
<th>Source</th>
<th>Two-tag</th>
<th>Three-tag</th>
<th>Four-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$G_{KK}$ Scalar</td>
<td>$G_{KK}$ Scalar</td>
<td>$G_{KK}$ Scalar</td>
</tr>
<tr>
<td>Luminosity</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>JER</td>
<td>0.25</td>
<td>0.74</td>
<td>1.4</td>
</tr>
<tr>
<td>JMR</td>
<td>0.52</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>JES/JMS</td>
<td>0.43</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>$b$-tagging</td>
<td>0.83</td>
<td>27</td>
<td>29</td>
</tr>
<tr>
<td>Bkgd estimate</td>
<td>2.8</td>
<td>-</td>
<td>5.8</td>
</tr>
<tr>
<td>Statistical</td>
<td>0.6</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Total Syst</td>
<td>3.1</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>31</td>
<td>32</td>
</tr>
</tbody>
</table>

An additional uncertainty in the normalization of the multijet background is derived from variations of the size or position of the control region or the sideband region. These variations shift the central position of the sideband and control regions by ±3 GeV in the directions of $m_{J}^{\text{lead}}$ or $m_{J}^{\text{subl}}$. The normalization fit to the leading large-$R$ jet’s mass distribution is carried out in the varied sideband region, and the validation is done in the varied control region. For each variation, the estimated total number of background events is compared to the data, and the largest difference seen is taken as the uncertainty. The assigned uncertainties are ±12.2%, ±4.2% and ±2.8% in four-tag, three-tag and two-tag samples, respectively.

Each uncertainty in the data-driven background estimate is evaluated for each sample, and these are treated as uncorrelated across the samples in the statistical analysis.

A summary of the systematic uncertainties and their impacts on the event yields is presented in table 6. The impact of $b$-tagging efficiency uncertainties is smaller in the three-tag sample, since the variations applied to $b$-tagged track-jets and anti-$b$-tagged track-jets have effects that partially cancel out.

6.4 Signal region event yields

The number of events observed in the data, the predicted number of background events in the signal region, and the predicted yield for two potential signals are presented in table 7 for the two-tag, three-tag, and four-tag samples. The scalar sample shown is normalized to 12 fb. The numbers of predicted background events and observed events are in agreement within the statistical uncertainties.

Figure 8 shows comparisons of the predicted $m_{2J}$ background distributions to those observed in the data. The predicted background and observed distributions are in agreement, with no significant excess.
Table 7. The number of predicted background events in the signal region for the boosted analysis compared to the data, for the two-tag, three-tag, and four-tag samples. The yields for a 2 TeV scalar and a 2 TeV $G_{KK}$ with $k/M_{Pl} = 1$ are also shown. The scalar is normalized to a cross section times branching ratio of 12 fb. The quoted uncertainties include both the statistical and systematic uncertainties. The anti-correlation between the multijet and $t\bar{t}$ yields is accounted for in the uncertainty in the total background yield.

<table>
<thead>
<tr>
<th></th>
<th>Two-tag</th>
<th>Three-tag</th>
<th>Four-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multijet</td>
<td>3390 ± 150</td>
<td>702 ± 63</td>
<td>32.9 ± 6.9</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>860 ± 110</td>
<td>80 ± 33</td>
<td>1.7 ± 1.4</td>
</tr>
<tr>
<td>Total</td>
<td>4250 ± 130</td>
<td>782 ± 51</td>
<td>34.6 ± 6.1</td>
</tr>
<tr>
<td>$G_{KK}$ (2 TeV)</td>
<td>0.97 ± 0.29</td>
<td>1.23 ± 0.16</td>
<td>0.40 ± 0.13</td>
</tr>
<tr>
<td>Scalar (2 TeV)</td>
<td>28.2 ± 9.0</td>
<td>35.0 ± 4.6</td>
<td>10.9 ± 3.5</td>
</tr>
<tr>
<td>Data</td>
<td>4376</td>
<td>801</td>
<td>31</td>
</tr>
</tbody>
</table>

Figure 8. Distributions of $m_{2J}$ in the signal regions of the boosted analysis for (a) the two-tag sample, (b) the three-tag sample, and (c) the four-tag sample, compared to the predicted backgrounds. The data-to-background ratio (bottom panels) shows also the combination of statistical and systematic uncertainties as the grey hatched band. The expected signal for a 2 TeV $G_{KK}$ resonance with $k/M_{Pl} = 1$ and a scalar with the same mass is also shown. The scalar has an arbitrary cross section times branching ratio of 12 fb.
7 Statistical analysis

Following the statistical procedures outlined in ref. [1], a test statistic based on the profile likelihood ratio [70] is used to test hypothesized values of $\mu = \sigma/\sigma_{\text{model}}$, the global signal strength factor, separately for each signal model. The exclusion limits are computed using asymptotic formulae [70] and are based on the $CL_s$ method [71], where a value of $\mu$ is regarded as excluded at the 95% confidence level (CL) when $CL_s$ is less than 5%.

7.1 Resonant $HH$ production

The resolved analysis is performed for resonance masses in the range 260–1400 GeV, the boosted analysis is carried out for signal masses in the range 800–3000 GeV, and the two analyses are combined in the mass range where they overlap. The statistical analysis is performed using the data observed in the signal regions. For the resolved analysis, the $m_{4j}$ distribution is used as the final discriminant and the 2015 and 2016 datasets are fitted simultaneously. In the boosted analysis the $m_{2J}$ distribution is used as the final discriminant, and the data from the two-tag, three-tag and four-tag signal regions are fitted simultaneously; however, the two-tag sample is not considered for masses below 1500 GeV since its contribution to the sensitivity is negligible at low mass.

Systematic uncertainties are treated within each signal region using Gaussian or log-normal constraint terms in the definition of the likelihood function. The luminosity uncertainty is treated as uncorrelated for the 2015 and 2016 datasets of the resolved analysis and the combined boosted dataset, since a subset of the 2016 dataset in the resolved analysis could not be used. All other systematic uncertainties affecting the signal are fully correlated between the resolved and boosted samples. Uncertainties in the background models are treated as uncorrelated between both analyses.

Before the fit is performed on the collision data, it is first validated on artificial datasets without statistical fluctuations. These datasets are created from the background-only model in the signal regions. The pulls, constraints, and correlations of the nuisance parameters are then checked at each mass point in the fits to the data. The impact of each uncertainty on $\mu$ is computed; the leading uncertainty at 280 GeV is the background modelling uncertainty that arises from comparing the background-model derived from control region data rather than sideband data. At 2000 GeV it is the background modelling uncertainty calculated by comparing smoothed data to the smoothed prediction in the control region.

Asymptotic approximations are used to calculate the local significance of a deviation from the background-only hypothesis. The largest local deviation is found at 280 GeV, where the $p_0$ value is $1.7 \cdot 10^{-4}$ (3.6$\sigma$) for the narrow-width scalar, and $5.8 \cdot 10^{-3}$ (2.5$\sigma$) for the $k/M_{\text{Pl}} = 1$ graviton model. The $k/M_{\text{Pl}} = 2$ graviton signal is too wide to explain this deviation. The signal shape of the scalar at 280 GeV has an approximately Gaussian form, resulting from the $m_{4j}$ resolution of about 9 GeV. The graviton signals have finite widths $\Gamma$ ($\Gamma = 8$ GeV for $k/M_{\text{Pl}} = 1$ and $\Gamma = 33$ GeV for $k/M_{\text{Pl}} = 2$) and furthermore, their shapes are distorted due to their close proximity to the kinematic threshold.

The global significance is evaluated using pseudo-experiments, generated from the background-only model that has been fitted to the data. For each pseudo-experiment, the
largest local excess is obtained from unconditional fits to all signal mass points and all three signal models. A distribution of those local significances is sampled, and the global $p_0$ value is computed by counting how often the largest excess in the pseudo-experiments is more significant than that observed in the data. The global significance obtained is 2.3$\sigma$.

Upper limits on the cross section are set in each of the benchmark models. Figure 9 shows the combined 95% CL upper limits for a narrow-width scalar and a spin-2 $G_{KK}$ in the bulk Randall-Sundrum (RS) model with $k/M_P = 1$ or $k/M_P = 2$. The predicted cross sections are shown for the graviton models, taken from ref. [36]. The bulk RS model with $k/M_P = 1$ is excluded for masses between 313 and 1362 GeV, and the bulk RS model with $k/M_P = 2$ is excluded for masses below 1744 GeV. The limits at low mass are stronger for the wider Graviton than the narrow Graviton model, because in the former case the signal has a larger acceptance and the distorted shape peaks at higher values of $m_{4j}$ above a steeply falling background.
The non-resonant search is performed using the resolved analysis only, since it has much better sensitivity to non-resonant signals than the boosted analysis. Using the SM $HH$ non-resonant production via gluon-gluon fusion as the signal model, computed at NLO and fully accounting for the top-quark mass, the observed 95% CL upper limit is $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}) < 147\text{ fb}$. This value is to be compared with the SM prediction for gluon-gluon fusion produced $HH$ of $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}) = 11.1^{+0.9}_{-1.0}\text{ fb}$. The expected and observed limit values and the uncertainties in the expectation are given in table 8, in units of the SM prediction. For ease of comparison with previous results, the limits are also calculated using the approximate top-quark mass correction: the observed limit is $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}) < 134\text{ fb}$, about 9% smaller than that obtained with the exact top-quark mass correction.

Despite the excess at 280 GeV, the non-resonant limits are stronger than expected. The non-resonant signal shape has its maximum around 400 GeV, with a slowly falling spectrum towards high mass (figure 4), and in several bins the number of observed events is below the prediction.

The result is limited by systematic uncertainties in the background normalization and shape. Since these are data-driven, an increase of the integrated luminosity will improve the sensitivity.

### 8 Conclusion

A search for both resonant and non-resonant production of pairs of Standard Model Higgs bosons has been carried out in the $b\bar{b}b\bar{b}$ channel, using 27.5–36.1 fb$^{-1}$ of LHC proton-proton collision data at $\sqrt{s} = 13\text{ TeV}$ collected by ATLAS in 2015 and 2016.

Results are reported for the resolved analysis with each $H \rightarrow b\bar{b}$ decay reconstructed as two separate $b$-tagged small-$R$ jets and for the boosted analysis with each $H \rightarrow b\bar{b}$ decay reconstructed as a single large-$R$ jet associated with at least one $b$-tagged track-jet.

No significant data excess is observed above the estimated background consisting mainly of multijet and $t\bar{t}$ events. The largest deviation from the background-only hypothesis is observed for narrow signal models at a mass of 280 GeV in the resolved analysis, with a global significance of 2.3$\sigma$.

Upper limits on the production cross section times branching ratio to the $b\bar{b}b\bar{b}$ final state are set for a narrow-width scalar and for spin-2 resonances with $k/M_{Pl} = 1$ or $k/M_{Pl} = 2$. The bulk RS model with $k/M_{Pl} = 1$ is excluded for masses between 313 and 1362 GeV, and the bulk RS model with $k/M_{Pl} = 2$ is excluded for masses below 1744 GeV. The 95% CL upper limit on the non-resonant production is 147 fb, which corresponds to 12.9 times the SM expectation.

<table>
<thead>
<tr>
<th>Observed</th>
<th>$-2\sigma$</th>
<th>$-1\sigma$</th>
<th>Expected</th>
<th>$+1\sigma$</th>
<th>$+2\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.9</td>
<td>11.1</td>
<td>14.9</td>
<td>20.7</td>
<td>30.0</td>
<td>43.6</td>
</tr>
</tbody>
</table>

Table 8. 95% CL exclusion limits for SM non-resonant $HH$ production, in units of the SM prediction for $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b})$. 

7.2 SM non-resonant $HH$ production
Acknowledgments

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

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNI SW and N CN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d’ Avenir Labex and Idex, 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 GIF, Israel; CERCA Programme Generalitat de Catalunya, Spain; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [72].

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[15] ATLAS collaboration, Searches for Higgs boson pair production in the $hh \to b\tau\tau$, $\gamma\gamma WW^*$, $\gamma\gamma bb$, $bbbb$ channels with the ATLAS detector, *Phys. Rev.* D 92 (2015) 092004 [arXiv:1509.04670] [SPIRE].


[29] M.H. Seymour and A. Siodmok, Constraining MPI models using $\sigma_{\text{HT}}$ and recent Tevatron and LHC underlying event data, JHEP 10 (2013) 113 [arXiv:1307.5015] [SPIRE].


The ATLAS collaboration

Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA; United States of America

Institut für Physik, Humboldt Universität zu Berlin, Berlin; Germany

Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern; Switzerland

School of Physics and Astronomy, University of Birmingham, Birmingham; United Kingdom

Centro de Investigaciones, Universidad Antonio Nariño, Bogota; Colombia

(a) Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna; (b) INFN Sezione di Bologna; Italy

Physikalisches Institut, Universität Bonn, Bonn; Germany

Department of Physics, Boston University, Boston MA; United States of America

Department of Physics, Brandeis University, Waltham MA; United States of America

(a) Transilvania University of Brașov, Brașov; (b) Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest; (c) Department of Physics, Alexandru Ioan Cuza University of Iași, Iași; (d) National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj-Napoca; (e) University Politehnica Bucharest, Bucharest; (f) West University in Timisoara, Timisoara; Romania

28 (a) Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava; (b) Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice; Slovak Republic

Physics Department, Brookhaven National Laboratory, Upton NY; United States of America

Departamento de Física, Universidad de Buenos Aires, Buenos Aires; Argentina

Cavendish Laboratory, University of Cambridge, Cambridge; United Kingdom

(a) Department of Physics, University of Cape Town, Cape Town; (b) Department of Mechanical Engineering Science, University of Johannesburg, Johannesburg; (c) School of Physics, University of the Witwatersrand, Johannesburg; South Africa

Department of Physics, Carleton University, Ottawa ON; Canada

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

CERN, Geneva; Switzerland

Enrico Fermi Institute, University of Chicago, Chicago IL; United States of America

LPC, Université Clermont Auvergne, CNRS/IN2P3, Clermont-Ferrand; France

Nevis Laboratory, Columbia University, Irvington NY; United States of America

Niels Bohr Institute, University of Copenhagen, Copenhagen; Denmark

(a) Dipartimento di Fisica, Università della Calabria, Rende; (b) INFN Gruppo Collegato di Cosenza, Laboratori Nazionali di Frascati; Italy

Physics Department, Southern Methodist University, Dallas TX; United States of America

Physics Department, University of Texas at Dallas, Richardson TX; United States of America

(a) Department of Physics, Stockholm University; (b) Oskar Klein Centre, Stockholm; Sweden

Deutsches Elektronen-Synchrotron DESY, Hamburg and Zeuthen; Germany

Lehrstuhl für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund; Germany

Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden; Germany

Department of Physics, Duke University, Durham NC; United States of America

SUPA — School of Physics and Astronomy, University of Edinburgh, Edinburgh; United Kingdom

INFN e Laboratori Nazionali di Frascati, Frascati; Italy

Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany

H. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen; Germany

Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève; Switzerland

(a) Dipartimento di Fisica, Università di Genova, Genova; (b) INFN Sezione di Genova; Italy