Search for a pseudoscalar boson in the mass range from 4 to 15 GeV produced in decays of the 125 GeV Higgs boson in the final states with two muons and two nearby tracks at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search is presented for pairs of light pseudoscalar bosons, in the mass range between 4 and 15 GeV, produced in decay of the 125 GeV Higgs boson. The decay mode where one light boson decays into $\tau$ leptons, while the other one decays into a pair of $\tau$ leptons or muons, is considered. The search is based on proton-proton collision data collected by the CMS experiment at a centre-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 35.9 fb$^{-1}$. No significant excess of events above the standard model background expectation obtained from control regions in data is observed. The 95% confidence level observed (expected) upper limits on the signal production cross section times branching fraction into the 4$\tau$ final state, relative to the standard model Higgs boson production cross section are set between 0.022 (0.027) and 0.23 (0.19) in the probed mass range.
1 Introduction

Searches for additional Higgs bosons, predicted by theories extending the standard model (SM), constitute important part of the scientific program at the CERN Large Hadron Collider (LHC). The analysis presented in this paper targets theoretical models with two Higgs doublets and one additional Higgs singlet complex field, which does not couple at tree level to fermions or gauge bosons and interacts only with itself and the Higgs doublets [1–8]. The Higgs sector of these models (denoted hereafter as 2HD+1S), features seven physical states, namely three CP-even, two CP-odd, and two charged bosons. Such a Higgs sector is realized for example in the next-to-minimal supersymmetric models that solve the so-called $\mu$-problem of the minimal supersymmetric extension of the SM [9]. A vast set of the 2HD+1S models is consistent with SM measurements and constraints from searches for additional Higgs bosons and supersymmetric particles, as well as with the measured properties of the Higgs boson [10–15].

The experimental study discussed in this note addresses a specific scenario of the 2HD+1S models in which the lightest pseudoscalar boson ($a_1$) with mass $2m_{a_1} < 125$ GeV has large singlet component, and therefore its couplings to the SM particles are significantly reduced. For this reason analyses making use of conventional production modes of $a_1$, such as gluon-gluon fusion (ggF) or b-quark associated production, have limited sensitivity to the expected signal. The $a_1$ boson is nonetheless potentially accessible in the decay of the 125 GeV Higgs boson ($H(125)$). The $a_1$ states are produced in such decay pairwise, $H(125) \rightarrow a_1a_1$, and can be identified via their decay into a pair of fermions [16–23]. The results of the study of the $H(125)$ boson properties allow the branching fraction of $H(125)$ decay into non-SM particles to be as high as 34% [24], which can potentially accommodate the $H(125) \rightarrow a_1a_1$ decay with sufficiently high rate for its detection at the LHC.

Several searches for $H(125) \rightarrow a_1a_1$ decays have been performed by the ATLAS and CMS Collaborations using Run 1 and Run 2 LHC data, exploiting various decays modes of the $a_1$ boson and probing different ranges of its mass [25–36]. None of these searches found any significant deviation from the background expectation and upper limits on the signal production cross section times branching fraction and constraints on parameters of the 2HD+1S models have been set.

This note presents an updated search for the $a_1$ boson in the decay channels $H(125) \rightarrow a_1a_1 \rightarrow 4\tau, 2\mu 2\tau$, using a data set corresponding to an integrated luminosity of 35.9 fb$^{-1}$, collected with the CMS detector in 2016 at centre-of-mass energy of 13 TeV. The search is performed for light $a_1$ states covering a mass range of 4 to 15 GeV and employs special analysis strategy to select and identify Lorentz-boosted $\tau$ lepton pairs with overlapping decay products. The study thus complements recent CMS searches for the $H(125) \rightarrow a_1a_1$ decay performed with Run 2 data in the $2\mu 2\tau$ [28], $2\tau 2b$ [29] and $4\mu$ [36] final states, covering a mass range of $0.25 < m_{a_1} < 62.5$ GeV. In this mass range decay products of the $a_1$ boson are expected to have relatively large angular separation and conventional analysis techniques to identify isolated leptons are applicable.

The branching fraction of the $a_1 \rightarrow \tau\tau$ decay depends on the way Higgs doublets couple to fermions and on the parameter $\tan \beta$, the ratio of the vacuum expectation values of two Higgs doublets [37]. In type-II 2HD+1S models, where one Higgs doublet couples to up-type fermions while the other couples to down-type fermions, the $a_1 \rightarrow \tau\tau$ decay rate is enhanced at high values of $\tan \beta$. The branching fraction of this decay reaches value above 90% at $\tan \beta > 3$ and $2m_{a_1} < 2m_b$, where $m_b$ is the mass of the bottom quark. For higher values of $m_{a_1}$, the branching fraction drops to 5–6%, since the decay into a pair of bottom quarks becomes kinematically possible and overwhelms the decay into a pair of $\tau$ leptons. In some of the 2HD+1S models however the $a_1 \rightarrow \tau\tau$ decay may be dominant even above the $a_1 \rightarrow b\bar{b}$ decay threshold. This
scenario is realized, e.g. for $\tan \beta > 1$ in the type-III 2HD+1S models, where one Higgs doublet couples to charged leptons, whereas the other doublet couples to quarks.

In the analysis presented here, each $a_1$ boson decay is identified by the presence of a muon and only one additional charged particle. This approach targets two decay channels, $a_1 \rightarrow \mu\mu$ and $a_1 \rightarrow \tau\tau_{\text{one-prong}}$. Here $\tau_{\text{mu}}$ denotes the muonic $\tau$ lepton decay and $\tau_{\text{one-prong}}$ stands for its leptonic or one-prong hadronic decay mode. Given the large difference in mass between the $a_1$ and the H(125) states, the $a_1$ bosons will be produced highly Lorentz boosted, and their decay products are expected to be strongly collimated, resulting in a signature with two muons, each of which is accompanied by a nearby particle with opposite an charge. The search primarily targets the dominant ggF process, in which the H(125) state is produced with relatively small transverse momentum ($p_T$) and the $a_1$ states are nearly back-to-back in the transverse plane, with a large separation in the azimuthal angle $\phi$ (in radians) between the particles originating from one $a_1$ decay and the particles produced by the decay of the other $a_1$ boson. In the ggF process the H(125) state can be also produced with relatively high boost when a hard gluon is radiated from the initial-state gluons or the heavy quark loop. In this case the separation in $\phi$ is reduced, but the separation in pseudorapidity $\eta$ can still be large. The properties discussed above define the topology of the signal, as illustrated in Fig. 1. The analysis therefore searches for a signal in a sample of dimuon events with large angular separation between the muons, where each muon is accompanied by one nearby opposite-sign particle originating in the same $a_1$ decay. The additional requirement of same-sign (SS) muons in the event largely suppresses $t\bar{t}$ and Drell-Yan background processes. This requirement also facilitates the implementation of a dedicated SS dimuon trigger with relatively low thresholds and acceptable rate.

![Lorentz-boosted states](image)

Figure 1: Illustration of the signal topology. The H(125) decays into two $a_1$ bosons, where one decays into a pair of $\tau$ leptons, while the other one decays into a pair of muons or a pair of $\tau$ leptons. The analyzed final state consists of one muon and an oppositely charged track in each $a_1$ decay leg.

2 CMS detector

The central feature of the CMS detector is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Forward calorimeters extend the eta coverage provided by the barrel and endcap detectors. Muons are detected in
gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid.

Events of interest are selected using a two-tiered trigger system [38]. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a time interval of less than $4 \mu s$. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate below 1 kHz before data storage.

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [39].

3 Simulated samples

For the simulation of the dominant ggF production mode, the Monte Carlo (MC) event generators \textsc{Pythia} (v.8.212) [40] and \textsc{MadGraph5}\_amc@$\text{NLO}$ [41] are used in order to model the $H(125) \rightarrow a_1 a_1 \rightarrow 4\tau$ and $H(125) \rightarrow a_1 a_1 \rightarrow 2\mu 2\tau$ signal events, respectively. For both decay modes the $p_T$ distribution of the $H(125)$ boson emerging from ggF is reweighted with next-to-next-to-leading order (NNLO) $K$ factors obtained by the program \textsc{HQT} (v2.0) [42, 43] with NNLO NNPDF3.0 parton distribution functions [44], hereby taking into account the more precise spectrum calculated to NNLO with resummation to next-to-next-to-leading-logarithms (NNLL) order. Subdominant contribution from other production modes of $H(125)$, namely vector boson fusion process (VBF), vector boson associated production (VH) and top quark pair associated production (ttH) are estimated using \textsc{Pythia}.

For the background studies, diboson and QCD multijet backgrounds are simulated with \textsc{Pythia}. Inclusive Z and W boson production processes are generated with \textsc{MadGraph5}\_amc@$\text{NLO}$ (v.2.2.2) generator at LO with the MLM jet matching and merging. The single-top and $t\bar{t}$ production are generated at NLO with the \textsc{Powheg} (v.2.0) generator. The set of parton distribution functions (pdf) is NLO NNPDF3.0 for NLO samples, and LO NNPDF3.0 for LO samples [44].

Showering and hadronization are carried out by \textsc{Pythia} with the CUETP8M1 underlying event tune [45], while a detailed simulation of the CMS detector is based on the \textsc{Geant4} [46] package.

4 Event selection

Events are selected using a SS dimuon trigger with thresholds on the $p_T$ of 17 (8) GeV for the leading (subleading) muon. To pass the high-level trigger, the tracks of the two muons are additionally required to have points of closest approach to the beam axis within 2 mm of each other along the longitudinal direction. Events recorded with this trigger are then reconstructed with the particle-flow (PF) algorithm [47]. The proton-proton (pp) interaction vertices are reconstructed using a Kalman filtering technique [48, 49]. Usually more than one such vertex is reconstructed, due to pileup, i.e., multiple pp collisions within the same or neighbouring bunch crossings, while the mean number of interactions per bunch crossing was 23 in 2016. The reconstructed vertex with the largest value of summed physics-object $p_T^2$ is taken to be the primary interaction vertex (PV). The physics objects are the jets, clustered using the jet finding algorithm [50, 51] with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the $p_T$ of those jets.

Events must contain at least two SS muons reconstructed with the PF algorithm [47], and each pair must fulfil the following requirements:
• The pseudorapidity of the leading and the subleading muons must be $|\eta| < 2.4$
• The $p_T$ of the leading (subleading) muon must exceed 18 (10) GeV
• The transverse (longitudinal) impact parameters of muons with respect to the PV are required to be $|d_0| < 0.05$ ($|dz| < 0.1$) cm
• The angular separation between the muons is required to be $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} > 2$, and if more than one SS muon pair is found in the event, the pair with the largest scalar sum of their muon transverse momenta is chosen

In the next step, the analysis employs information about tracks associated with the reconstructed charged PF objects, excluding the pair of SS muons identified and selected as described above. Each selected muon, is required to have tracks with $p_T > 1$ GeV and $|\eta| < 2.4$, while both impact parameters of $|d_0|$ and $|dz|$ have to be less than 1 cm, within a cone of size $\Delta R = 0.5$ centered on the muon momentum direction. The loose impact parameter requirements on the tracks are designed to suppress background events in which a heavy-flavor hadron decays into a muon and several charged particles. Although tracks from these decay products are expected to be displaced from the PV, they can still satisfy the loose track impact parameter criteria. The contribution from such events to the final selected sample is further suppressed by the requirement of exactly one track accompanying the muon.

The selected events are required to have exactly two $a_1 \to \tau_\mu \tau_\tau$ one-prong pairs, and for each one, only one track must be within a cone of size $\Delta R = 0.5$ around the muon, and it has to satisfy the following selection criteria:
• The track charge must be opposite to that of the nearby muon.
• The track must have $p_T > 2.5$ GeV and $|\eta| < 2.4$.
• The transverse and longitudinal impact parameters of the track are required to be $|d_0| < 0.02$ cm and $|dz| < 0.04$ cm, respectively.

The set of selection requirements outlined above defines the signal region (SR).

The expected signal yield and signal acceptance for few representative values of $m_{a_1}$ are reported in Table 1. The signal yields are computed for the benchmark value of the branching fraction, $B(H(125) \to a_1a_1) \cdot B^2(a_1 \to \tau \tau) = 0.2$ and assuming that the H(125) production cross section is the one predicted in the SM. The yield of the ggF signal events followed by the $H(125) \to a_1a_1 \to 2\mu 2\tau$ decay is estimated under assumption that partial widths of the $a_1 \to \mu \mu$ and $a_1 \to \tau \tau$ decays satisfy the relation [52]. Under the same assumption the contribution from the $H(125) \to a_1a_1 \to 4\mu$ decay is evaluated to range between 0.7 and 3%, depending on the probed mass of the $a_1$ boson. This contribution is not considered in the present analysis.

$$\frac{\Gamma(a_1 \to \mu \mu)}{\Gamma(a_1 \to \tau \tau)} = \frac{m_\mu^2}{m_\mu^2 \sqrt{1 - (2m_\tau/m_{a_1})^2}}. \tag{1}$$

The ratio of branching fractions of the $a_1a_1 \to 2\mu 2\tau$ and $a_1a_1 \to 4\tau$ decays is computed through the ratio of the partial widths $\Gamma(a_1 \to \mu \mu)$ and $\Gamma(a_1 \to \tau \tau)$ as

$$\frac{B(a_1a_1 \to 2\mu 2\tau)}{B(a_1a_1 \to 4\tau)} = \frac{2 \Gamma(a_1 \to \mu \mu)}{\Gamma(a_1 \to \tau \tau)} = \frac{2}{2} \frac{\Gamma(a_1 \to \mu \mu)}{\Gamma(a_1 \to \tau \tau)}. \tag{2}$$
The factor of 2 in Eq. (2) arises from two possible decays, \( a_1^{(1)} a_1^{(2)} \rightarrow 2\mu 2\tau \) and \( a_1^{(1)} a_1^{(2)} \rightarrow 2\tau 2\mu \), that produce the final state with two muons and two \( \tau \) leptons. The ratio in Eq. 2 ranges from about 0.0073 at \( m_{a_1} = 15 \text{ GeV} \) to about 0.0155 at \( m_{a_1} = 4 \text{ GeV} \).

A simulation-based study shows that the QCD multijet events dominate the final selected sample. Contribution from other backgrounds sources constitutes about 1% of all selected events.

Table 1: The signal acceptance and the number of expected signal events after final selection. The former is presented for the final state with two same-sign muons, while the signal yields are shown for the benchmark value of branching fraction, \( B(H(125) \rightarrow a_1 a_1) \cdot B^2(a_1 \rightarrow \tau \tau) = 0.2 \). The quoted uncertainties for predictions from simulation include only statistical ones. The number of observed events selected in the signal region amounts to 2035.

<table>
<thead>
<tr>
<th>( m_{a_1} ) [GeV]</th>
<th>Acceptance ( \times 10^3 )</th>
<th>( 4\tau )</th>
<th>( 2\mu 2\tau )</th>
<th>( 4\tau )</th>
<th>( 2\mu 2\tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6.51 ± 0.31</td>
<td>28.10 ± 0.45</td>
<td>129.9 ± 6.2</td>
<td>54.7 ± 0.9</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.95 ± 0.28</td>
<td>21.72 ± 0.45</td>
<td>98.8 ± 5.5</td>
<td>22.5 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.90 ± 0.21</td>
<td>14.82 ± 0.38</td>
<td>57.8 ± 4.2</td>
<td>14.2 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.43 ± 0.08</td>
<td>1.09 ± 0.10</td>
<td>8.5 ± 1.1</td>
<td>1.0 ± 0.1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Binning of the 2D \((m_1, m_2)\) distribution. The last bin is an overflow bin and includes also masses up to 15 GeV, while values in parentheses are given in GeV.

5 Modelling of the background

A simulation-based study reveals that the sample of SS muon pairs selected as described in Section 4, but without requiring a presence of one-prong \( \tau \) candidates and without applying the isolation requirement for the muon-track systems, is dominated by QCD multijet events, where about 85% of all selected events contain bottom quarks in the final state. The same-sign muon pairs in these events originate mainly from the following sources:

- muonic decay of a bottom hadron in one bottom quark jet and cascade decay of a bottom hadron into a charm hadron with a subsequent decay in the other bottom quark jet;
• muonic decay of a bottom hadron in one bottom quark jet and decay of a quarkonium state into a pair of muons in the other jet;
• muonic decay of a bottom hadron in one bottom quark jet and muonic decay of a $B^0$ meson in the other bottom quark jet. The SS muon pair in this case may appear as a result of $B^0$–$\bar{B}^0$ oscillations.

The two-dimensional (2D) distribution of the invariant masses of the muon-track systems is used to discriminate between signal and the dominant QCD multijet background in the signal extraction procedure. The 2D distribution is filled with a pair of the muon-track invariant masses $(m_1, m_2)$, ordered by their value, $m_2 > m_1$. The binning of the 2D distribution adopted in the analysis is illustrated in Fig. 2, while the last bin been an overflow bin including all masses up to 15 GeV. As $m_2$ is required to exceed $m_1$, only $(i, j)$ bins with $j \geq i$ are filled in the 2D distribution, yielding in total $N \times (N + 1)/2 = 6 \times (6 + 1)/2 = 21$ independent bins. Bins $(i, 6)$ with $i=1, 5$ contain all events with $m_2 > 6$ GeV. Bin $(6, 6)$ contains all events with $m_{1,2} > 6$ GeV.

The normalization of the background is not constrained prior to the extraction of the signal. The procedure used to model the shape of the (2D) $(m_1, m_2)$ distribution of background events is described in this section.

The normalized 2D $(m_1, m_2)$ distribution for the muon-track pairs with $m_2 > m_1$ is represented by a binned template constructed using the following relation

$$f_{2D}(i, j) = C(i, j)(f_{1D}(i) \times f_{1D}(j))^{sym},$$

with

$$(f_{1D}(i) \times f_{1D}(i))^{sym} = f_{1D}(i) \cdot f_{1D}(i),$$

$$f_{1D}(i) \times f_{1D}(j) = f_{1D}(i) \cdot f_{1D}(j) + f_{1D}(j) \cdot f_{1D}(i), \quad \text{if } j > i,$$

where

• $f_{2D}(i, j)$ is the content of the bin $(i, j)$ in the normalized 2D $(m_1, m_2)$ distribution;
• $f_{1D}(i)$ is the content of bin $i$ in the normalized one-dimensional (1D) distribution of the muon-track invariant mass;
• $C(i, j)$ is a symmetric matrix, accounting for possible correlation between $m_1$ and $m_2$.

The elements of the matrix $C(i, j)$ are referred to as “correlation factors” in the following. The condition $C(i, j) = 1$ for all bins $(i, j)$ would indicate an absence of correlation between $m_1$ and $m_2$. We sum the contents of the nondiagonal bins $(i, j)$ and $(j, i)$ in the Cartesian product of the 1D distributions $f_{1D}(i) \times f_{1D}(j)$ to account for the fact that each event enters the 2D $(m_1, m_2)$ distribution with ordered values of the muon-track invariant masses.

By construction the background model estimates the dominant QCD multijet production as well as small contributions from other processes.

Multiple control regions (CRs) are introduced in order to derive and validate the modelling of $f_{1D}(i)$ and $C(i, j)$. The CRs are defined on the basis of the modified isolation criteria applied to one or both muon-track pairs. The isolation criteria are specified by the multiplicity of tracks with $p_T > 1$ GeV and $|\eta| < 2.4$ in the cone of $\Delta R = 0.5$ around the muon momentum direction. The summary of all CRs used to derive and validate the modelling of background shape is...
5. Modelling of the background

Table 2: The list of control regions used to model background shape. $N_{\text{trk}}$ denotes the number of close-by tracks within a cone of $\Delta R = 0.5$ around the muon momentum direction. $N_{\text{trk,soft}}$ is the number of soft tracks with $1.0 \text{ GeV} < p_T < 2.5 \text{ GeV}$ in the cone of $\Delta R = 0.5$ around the muon momentum direction.

<table>
<thead>
<tr>
<th>Control region</th>
<th>Tag muon-trk</th>
<th>Probe muon-trk</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{trk}} = 1$</td>
<td>isolated</td>
<td>$N_{\text{trk}} &gt; 1$</td>
<td>validation of $f_{1\text{D}}(i)$</td>
</tr>
<tr>
<td>$N_{\text{trk}} = 2, 3$</td>
<td>$N_{\text{trk}} = 2, 3$</td>
<td>$N_{\text{trk}} &gt; 1$</td>
<td>validation of $f_{1\text{D}}(i)$</td>
</tr>
<tr>
<td>$N_{23}$</td>
<td>$N_{\text{trk}} = 2, 3$</td>
<td>isolated</td>
<td>determination of $f_{1\text{D}}(i)$</td>
</tr>
<tr>
<td>$N_{56}$</td>
<td>$N_{\text{trk}} = 4, 5$</td>
<td>isolated</td>
<td>systematics of $f_{1\text{D}}(i)$</td>
</tr>
<tr>
<td>both muon-track systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loose-Iso</td>
<td>$N_{\text{trk,soft}} = 1, 2$</td>
<td></td>
<td>determination of $C(i,j)$</td>
</tr>
</tbody>
</table>

given in Table 2.

5.1 Modelling of $f_{1\text{D}}(i)$

The normalized one-dimensional distribution $f_{1\text{D}}(i)$ is modeled using the $N_{23}$ CR. Events in this CR are required to satisfy all nominal selection criteria, except for the isolation of the second muon-track system. The second muon is required to be accompanied by either two or three nearby tracks with $p_T > 1 \text{ GeV}$ and impact parameters smaller than 1 cm relative to the PV, both in the transverse plane and along the beam axis. The simulation shows that more than 95% of events selected in the CR $N_{23}$ are QCD multijet events. The modelling of the $f_{1\text{D}}(i)$ template is based on the hypothesis that the kinematic distributions for the first muon-track system are weakly affected by the isolation requirement imposed on the second; therefore the $f_{1\text{D}}(i)$ distribution of the isolated muon-track system is expected to be similar in the SR and the $N_{23}$ CR.

This hypothesis is verified in control regions labelled $N_{\text{trk}} = 1$ and $N_{\text{trk}} = 2, 3$. Events are selected in these CR if one of the muons has at least one track passing the $\tau_{\text{one-prong}}$ candidate criteria within a $\Delta R$ cone of radius 0.5 around the muon direction, with any number of additional tracks within the same $\Delta R$ cone. As more than one of these tracks can pass this $\tau_{\text{one-prong}}$ criteria, two scenarios, namely the lowest and highest $p_T$ (“softest” and “hardest”) tracks passing the $\tau_{\text{one-prong}}$ decay candidate were investigated. In both scenarios, the corresponding track passing the one-prong decay candidate criteria is used to calculate the muon-track invariant mass, while if only one $\tau_{\text{one-prong}}$ candidate is found around the first muon, the track is regarded as both “hardest” and “softest”. For the second muon, two isolation requirements are considered: when the muon is accompanied by only one track passing the $\tau_{\text{one-prong}}$ decay candidate criteria (CR $N_{\text{trk}} = 1$) as in the SR, or when it is accompanied by two or three tracks ($N_{\text{trk}} = 2, 3$ CR) with $p_T > 1 \text{ GeV}$ and impact parameters smaller than 1 cm relative to the PV as in the region $N_{23}$ CR. The shapes of invariant mass distributions of the first muon and the softest or hardest accompanying track are then compared for the two different isolation requirements on the second muon, $N_{\text{trk,2}} = 1$ and $N_{\text{trk,2}} = 2, 3$. The results of this study are illustrated in Fig. 3. In both considered cases, the shape of the invariant mass distribution is compatible within statistical uncertainties between the two cases, $N_{\text{trk,2}} = 1$ and $N_{\text{trk,2}} = 2, 3$. This observation validates the assumption that the $f_{1\text{D}}(i)$ can be determined from the $N_{23}$ CR.

The potential dependence of the muon-track invariant mass distribution on the isolation requirement imposed on the second muon-track pair is verified also by comparing shapes in the control regions $N_{23}$ and $N_{45}$. The latter CR is defined by requiring the presence of 4 or 5
tracks around the second muon, while the first muon-track pair passes selection criteria for the \( a_1 \rightarrow \tau_\mu \tau_\mu \) candidate. The results are illustrated in Fig. 4. A slight difference is observed between distributions in these two CRs. This difference is taken as a shape uncertainty in the normalized 1D template \( f_{1D}(j) \) entering Eq. (3).

Fig. 5 presents the normalized invariant mass distribution of the muon-track system for data selected in the SR and for the background model derived from the \( N_{23} \) CR. The data and background distributions are compared to the signal distribution normalized to unity (signal pdf), obtained from simulation, for four representative mass hypotheses, \( m_{a_1} = 4, 7, 10, \) and \( 15 \) GeV. The invariant mass of the muon-track system is found to have higher discrimination power between the background and the signal at higher \( m_{a_1} \). For lower masses, the signal shape becomes more background like, resulting in a reduction of discrimination power.

### 5.2 modelling of \( C(i,j) \)

In order to determine the correlation factors \( C(i,j) \), an additional CR, labelled Loose-Iso, is used. It consists of events that contain two SS muons passing the identification and kinematic selection criteria outlined in Section 4. Each muon is required to have two or three nearby tracks within \( \Delta R < 0.5 \) around the muon direction. Only one of these tracks must satisfy the criteria imposed on the \( \tau_{\text{one-prong}} \) lepton decay candidates with \( p_T > 2.5 \) GeV, while the additional tracks must have \( 1.0 < p_T < 2.5 \) GeV. About 36k data events are selected in this CR. The simulation predicts that the QCD multijet events dominate this CR, comprising more than 99% of selected events. It was also found that the overall background-to-signal ratio is enhanced compared to the SR by a factor of 30 to 40, depending on the mass hypothesis of \( m_{a_1} \). The potential signal contamination in individual bins of the mass distributions has been investigated and does not exceed 3% for all bins of the 2D distributions and for all probed \( m_{a_1} \) masses. The event sample in this region is used to build the normalized 2D distribution.
5. Modelling of the background

Figure 4: The observed invariant mass distribution, normalized to unity, of the muon-track invariant mass in control regions $N_{23}$ (circles) and $N_{45}$ (squares).

$(f_{2D}(i,j))$. Finally, the correlation factors $C(i,j)$ are obtained according to Eq. (3) as:

$$C(i,j) = \frac{f_{2D}(i,j)}{(f_{1D}(i) \times f_{1D}(j))^{\text{sym}}},$$

(4)

where $f_{1D}(i)$ is the 1D normalized distribution with two entries per event ($m_1$ and $m_2$). The correlation factors $C(i,j)$ derived from data in the Loose-Iso CR are presented in Fig. 6. To obtain estimates of $C(i,j)$ in the signal region, the measurement in the Loose-Iso CR is corrected for the difference in $C(i,j)$ between the signal region and Loose-Iso CR. This difference is assessed with the samples of simulated background events. The correlation factors estimated from simulation in the signal region and the Loose-Iso CR are presented in Fig. 7.

The correlation factors in the signal region are then computed as

$$C(i,j) = C(i,j)_{\text{data}}^{\text{CR}} \frac{C(i,j)^{\text{sig}}_{\text{MC}}}{C(i,j)^{\text{CR}}_{\text{MC}}}$$

(5)

where

- $C(i,j)^{\text{CR}}_{\text{data}}$ are correlation factors derived for the Loose-Iso CR in data (Fig. 6);
- $C(i,j)^{\text{sig}}_{\text{MC}}$ are correlation factors derived for the SR in the simulated QCD multijet sample (Fig. 7, left);
- $C(i,j)^{\text{CR}}_{\text{MC}}$ are correlation factors derived for the Loose-Iso CR in the simulated QCD multijet sample (Fig. 7, right).

The difference in correlation factors derived in the SR and in the Loose-Iso CR is taken into account as an uncertainty in $C(i,j)$. 

---

5. Modelling of the background

Figure 4: The observed invariant mass distribution, normalized to unity, of the muon-track invariant mass in control regions $N_{23}$ (circles) and $N_{45}$ (squares).

$(f_{2D}(i,j))$. Finally, the correlation factors $C(i,j)$ are obtained according to Eq. (3) as:

$$C(i,j) = \frac{f_{2D}(i,j)}{(f_{1D}(i) \times f_{1D}(j))^{\text{sym}}},$$

(4)

where $f_{1D}(i)$ is the 1D normalized distribution with two entries per event ($m_1$ and $m_2$). The correlation factors $C(i,j)$ derived from data in the Loose-Iso CR are presented in Fig. 6. To obtain estimates of $C(i,j)$ in the signal region, the measurement in the Loose-Iso CR is corrected for the difference in $C(i,j)$ between the signal region and Loose-Iso CR. This difference is assessed with the samples of simulated background events. The correlation factors estimated from simulation in the signal region and the Loose-Iso CR are presented in Fig. 7.

The correlation factors in the signal region are then computed as

$$C(i,j) = C(i,j)_{\text{data}}^{\text{CR}} \frac{C(i,j)^{\text{sig}}_{\text{MC}}}{C(i,j)^{\text{CR}}_{\text{MC}}}$$

(5)

where

- $C(i,j)^{\text{CR}}_{\text{data}}$ are correlation factors derived for the Loose-Iso CR in data (Fig. 6);
- $C(i,j)^{\text{sig}}_{\text{MC}}$ are correlation factors derived for the SR in the simulated QCD multijet sample (Fig. 7, left);
- $C(i,j)^{\text{CR}}_{\text{MC}}$ are correlation factors derived for the Loose-Iso CR in the simulated QCD multijet sample (Fig. 7, right).

The difference in correlation factors derived in the SR and in the Loose-Iso CR is taken into account as an uncertainty in $C(i,j)$. 

---

5. Modelling of the background

Figure 4: The observed invariant mass distribution, normalized to unity, of the muon-track invariant mass in control regions $N_{23}$ (circles) and $N_{45}$ (squares).

$(f_{2D}(i,j))$. Finally, the correlation factors $C(i,j)$ are obtained according to Eq. (3) as:

$$C(i,j) = \frac{f_{2D}(i,j)}{(f_{1D}(i) \times f_{1D}(j))^{\text{sym}}},$$

(4)

where $f_{1D}(i)$ is the 1D normalized distribution with two entries per event ($m_1$ and $m_2$). The correlation factors $C(i,j)$ derived from data in the Loose-Iso CR are presented in Fig. 6. To obtain estimates of $C(i,j)$ in the signal region, the measurement in the Loose-Iso CR is corrected for the difference in $C(i,j)$ between the signal region and Loose-Iso CR. This difference is assessed with the samples of simulated background events. The correlation factors estimated from simulation in the signal region and the Loose-Iso CR are presented in Fig. 7.

The correlation factors in the signal region are then computed as

$$C(i,j) = C(i,j)_{\text{data}}^{\text{CR}} \frac{C(i,j)^{\text{sig}}_{\text{MC}}}{C(i,j)^{\text{CR}}_{\text{MC}}}$$

(5)

where

- $C(i,j)^{\text{CR}}_{\text{data}}$ are correlation factors derived for the Loose-Iso CR in data (Fig. 6);
- $C(i,j)^{\text{sig}}_{\text{MC}}$ are correlation factors derived for the SR in the simulated QCD multijet sample (Fig. 7, left);
- $C(i,j)^{\text{CR}}_{\text{MC}}$ are correlation factors derived for the Loose-Iso CR in the simulated QCD multijet sample (Fig. 7, right).

The difference in correlation factors derived in the SR and in the Loose-Iso CR is taken into account as an uncertainty in $C(i,j)$.
Figure 5: Normalized invariant mass distribution of the muon-track system for events passing the signal selection. Observed numbers of events are represented by data points with error bars. The QCD multijet background model is derived from the control region $N_{23}$. Also shown are the normalized distributions from signal simulations for four mass hypotheses, $m_a = 4, 7, 10,$ and $15$ GeV (dashed histograms). Each event in the observed and expected signal distributions contributes two entries, corresponding to the two muon-track systems in each event passing the requirements. The lower panel shows the ratio of the observed to expected number of background events in each bin of the distribution. The grey shaded area represents the statistical uncertainties.
5. Modelling of the background

Figure 6: The \((m_1, m_2)\) correlation factors \(C(i, j)\) with their statistical uncertainties, derived from data in the CR Loose-Iso.

Figure 7: Left (Right): The \((m_1, m_2)\) correlation factors \(C(i, j)\) along with their MC statistical uncertainties, derived from simulated samples in the signal region (Loose-Iso CR).
6 Signal modelling

The signal templates are derived from the simulated samples of the $H(125) \rightarrow a_1a_1 \rightarrow 4\tau$ and $H(125) \rightarrow a_1a_1 \rightarrow 2\mu2\tau$ decays. The study probes the signal strength modifier, defined as the ratio of the measured signal cross section times the branching fraction $B(H(125) \rightarrow a_1a_1)B^2(a_1 \rightarrow \tau\tau)$ to the inclusive cross section of the $H(125)$ boson production predicted in the SM. The relative contributions from different production modes of $H(125)$ are defined by the corresponding cross sections predicted in the SM. The relative contribution of the ggF production followed by the $H(125) \rightarrow a_1a_1 \rightarrow 2\mu2\tau$ decay, is computed assuming that the partial widths of $a_1 \rightarrow \tau\tau$ and $a_1 \rightarrow \mu\mu$ decays satisfy Eq. (1).

The invariant mass distribution of the muon-track system in the $a_1 \rightarrow \mu\mu$ decay channel exhibits resonance structure, peaking at the nominal value of the $a_1$ boson mass. Furthermore, the reconstructed muon-track mass in the $a_1 \rightarrow \tau\tau$ decay is typically lower than in the $a_1 \rightarrow \mu\mu$ decay. This is why the $H(125) \rightarrow a_1a_1 \rightarrow 2\mu2\tau$ signal samples have a largely different shape of the 2D $(m_1, m_2)$ distribution compared to the $H(125) \rightarrow a_1a_1 \rightarrow 4\tau$ signal samples. Fig. 8 compares the $(m_1, m_2)$ shown in a one-row distribution between the $H(125) \rightarrow a_1a_1 \rightarrow 4\tau$ and $H(125) \rightarrow a_1a_1 \rightarrow 2\mu2\tau$ signal samples for mass hypotheses $m_{a_1} = 4$ GeV and 10 GeV. The signal distributions are normalized assuming the SM production rate of the $H(125)$ boson as the one in the SM and setting the branching fraction $B(H(125) \rightarrow a_1a_1)B^2(a_1 \rightarrow \tau\tau) = 0.2$.

![Figure 8: The distribution of the signal templates $f_{2D}(i, j)$ in one row for mass hypotheses $m_{a_1} = 4$ GeV (left plot) and 10 GeV (right plot). The $H(125) \rightarrow a_1a_1 \rightarrow 2\mu2\tau$ (blue histogram) and $H(125) \rightarrow a_1a_1 \rightarrow 4\tau$ (red histogram) contributions are shown.](image)

7 Systematic uncertainties

The systematic uncertainties are grouped into two categories. The first one consists of uncertainties related to the background, while the second includes those related to the signal. All of the systematic uncertainties are summarized in Table 3.

7.1 Uncertainties related to the background

The estimation of the QCD multijet background is based on observed data, therefore it is not affected by imperfections in the simulation, reconstruction, or detector response.

The shape of the background in the 2D $(m_1, m_2)$ distribution is modeled according to Eq. (3), while its uncertainty is dominated by uncertainties related to the correlation factors $C(i, j)$ (as described in Section 5.2). Additionally, it is also affected by the shape uncertainty in the 1D
Table 3: Systematic uncertainties and their effect on the estimates of the QCD multijet background and signal. The effect of the uncertainties in $C(i,j)$ on the total background yield is absorbed by the overall background normalization, which is allowed to vary freely in the fit.

<table>
<thead>
<tr>
<th>Source</th>
<th>Value</th>
<th>Affected sample</th>
<th>Type</th>
<th>Effect on the total yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical uncertainties in $C(i,j)$</td>
<td>3–60%</td>
<td>bkg.</td>
<td>bin-by-bin</td>
<td>–</td>
</tr>
<tr>
<td>Extrapolation uncertainties in $C(i,j)$</td>
<td>–</td>
<td>bkg.</td>
<td>shape</td>
<td>–</td>
</tr>
<tr>
<td>Uncertainty in the 1D template $f_{1D}(i)$</td>
<td>–</td>
<td>bkg.</td>
<td>shape</td>
<td>–</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>2.5%</td>
<td>signal</td>
<td>norm.</td>
<td>2.5%</td>
</tr>
<tr>
<td>Muon identification and trigger efficiency</td>
<td>2% per muon</td>
<td>signal</td>
<td>norm.</td>
<td>4%</td>
</tr>
<tr>
<td>Track selection and isolation efficiency</td>
<td>4–12% per track</td>
<td>signal</td>
<td>shape</td>
<td>10–18%</td>
</tr>
<tr>
<td>MC stat. uncertainties in signal yields</td>
<td>8–100%</td>
<td>signal</td>
<td>bin-by-bin</td>
<td>5–20%</td>
</tr>
<tr>
<td>Theory uncertainties in the signal acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_R$ and $\mu_F$ variations</td>
<td></td>
<td></td>
<td></td>
<td>$&lt; 2%$</td>
</tr>
<tr>
<td>PDF (VBF, VH, t(\bar{t})H)</td>
<td></td>
<td></td>
<td></td>
<td>$2%$</td>
</tr>
<tr>
<td>Theory uncertainties in the signal cross sections</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_R$ and $\mu_F$ variations (gg → H(125))</td>
<td></td>
<td></td>
<td>norm.</td>
<td>$+3.8%$ $-6.7%$</td>
</tr>
<tr>
<td>$\mu_R$ and $\mu_F$ variations (VBF)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>$+0.4%$ $-0.3%$</td>
</tr>
<tr>
<td>$\mu_R$ and $\mu_F$ variations (VH)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>$+1.8%$ $-1.6%$</td>
</tr>
<tr>
<td>$\mu_R$ and $\mu_F$ variations (t(\bar{t})H)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>$+5.8%$ $-9.2%$</td>
</tr>
<tr>
<td>PDF (ggF → H(125))</td>
<td></td>
<td></td>
<td>norm.</td>
<td>3.1%</td>
</tr>
<tr>
<td>PDF (VBF)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>2.1%</td>
</tr>
<tr>
<td>PDF (VH)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>1.8%</td>
</tr>
<tr>
<td>PDF (t(\bar{t})H)</td>
<td></td>
<td></td>
<td>norm.</td>
<td>3.6%</td>
</tr>
</tbody>
</table>
template $f_{1D}(m)$ (as discussed in Section 5.1). The bin-by-bin uncertainties in mass correlation factors $C(i,j)$, derived from Eq. (5), are composed of the statistical uncertainties in observed data and simulated samples, as presented in Figs. 6-7, and range from 3 to 60%. These uncertainties are accounted for in the signal extraction procedure by one nuisance parameters per bin in the $(m_1,m_2)$ distribution. The systematic uncertainties related to the extrapolation of $C(i,j)$ from the Loose-Iso CR to the SR are derived from the dedicated MC study outlined in Section 5.2. The related shape uncertainty is determined by comparing correlation factors derived in the simulated samples, between the signal region and the Loose-Iso CR.

The impact of possible signal contamination in the Loose-Iso CR is estimated on a bin-by-bin basis, and it is at most 2.8% in the bin (6,6), when conservatively assuming that $B(H(125) \rightarrow a_1a_1) \cdot B^2(a_1 \rightarrow \tau\tau) = 0.34$, corresponding to an upper limit at 95% confidence level (CL) on the branching fraction of the H(125) decay into non-SM particles from Ref. [24]. The impact of this uncertainty is found to have negligible effect on the final results.

### 7.2 Uncertainties related to signal

An uncertainty of 2.5% is assigned to the integrated luminosity estimate [53]. The uncertainty in the muon identification and trigger efficiency is estimated to be 2% for each selected muon. The track selection and muon-track isolation efficiency is assessed with a study performed on a sample of Z bosons decaying into a pair of $\tau$ leptons. In the selected $Z \rightarrow \tau\tau$ events, one $\tau$ lepton is identified via its muonic decay, while the other is identified as an isolated track resulting from a one-prong decay. The track is required to pass the nominal selection criteria used in the main analysis. From this study, the uncertainty in the track selection and isolation efficiency is evaluated. The related uncertainty affects the shape of the signal estimate, while changing the overall signal yield by 10–18%. The muon and track momentum scale uncertainties are smaller than 0.3% and have a negligible effect on the analysis.

The bin-by-bin statistical uncertainties in the signal acceptance range from 8 to 100%, while the impact on the overall signal normalization varies between 5 and 20%.

Theoretical uncertainties have an impact on the differential kinematic distributions of the produced H(125) boson, in particular its $p_T$ spectrum, thereby affecting signal acceptance. The uncertainty due to missing higher-order corrections to the ggF process is estimated with the HQT program by varying the renormalization ($\mu_R$) and factorization ($\mu_F$) scales. The H(125) $p_T$-dependent $K$ factors are recomputed according to these variations and applied to the simulated signal samples. The resulting effect on the signal acceptance is estimated to vary between 1.2 and 1.5%, depending on $m_{a_1}$. In a similar way, the uncertainty in the signal acceptance is computed for the VBF, VH and $t\bar{t}H$ production processes. The impact on the acceptance is estimated to vary between 0.8 and 2.0%, depending on the process and probed mass of the $a_1$ boson.

The HQT program is also used to evaluate the effect of the PDF uncertainties. The nominal $K$ factors for the H(125) boson $p_T$ spectrum are computed with the NNPDF3.0 PDF set [44]. Variations of the NNPDF3.0 PDFs within their uncertainties change the signal acceptance by about 1%, whilst using the CTEQ6L1 PDF set [54] changes the signal acceptance by about 0.7%. The impact of the PDF uncertainties on the acceptance for the VBF, VH and $t\bar{t}H$ production processes is estimated in the same way and a 2% conservative uncertainty is considered to account for these.
8 Results

The signal is extracted with a binned maximum-likelihood fit applied to the 2D \((m_1, m_2)\) distribution. For each probed mass of the \(a_1\) boson, the \((m_1, m_2)\) distribution is fitted with the sum of two templates, corresponding to expectations for the signal and the overall background, dominated by QCD multijet events. The modelling of the background is discussed in Section 5. The construction of signal templates is described in Section 6.

The normalization of both signal and background are allowed to float freely in the fit. The systematic uncertainties, affecting normalization of signal templates, are incorporated in the fit via nuisance parameters with log-normal probability density function. The shape altering systematic uncertainties are represented by nuisance parameters, whose variations cause continuous morphing of the background template shape, and are assigned Gaussian probability density functions. The bin-by-bin statistical uncertainties are assigned gamma probability density functions.

Fig. 9 shows the distribution of \((m_1, m_2)\), where the notation for the bins follows that of Fig. 2. The distribution for the background is obtained by applying a fit to the observed data under the background-only hypothesis. Also shown are the expectations for the signal at \(m_{a_1} = 4, 7, 10, \) and \(15\) GeV. The signal normalization is computed assuming that the H(125) boson is produced in pp collisions with a rate predicted by the SM, and decays into \(a_1a_1 \rightarrow 4\tau\) final state with a branching fraction of 20\%. No significant deviations of data from the background expectation are observed in the 2D \((m_1, m_2)\) distribution.

Results of the analysis are used to set upper limits at 95\% CL on the signal cross section times the branching fraction, \(\sigma(pp \rightarrow H(125) + X) \cdot B(H(125) \rightarrow a_1a_1) \cdot B^2(a_1 \rightarrow \tau\tau)\), relative
to the inclusive SM cross section of $H(125)$ production. The modified frequentist CL$_s$ criterion [55, 56], and the asymptotic formulae are used for the test statistic [57], implemented in the ROOSTATS package [58]. Fig. 10 shows the observed and expected upper limits at 95% CL on the signal cross section times the branching fraction, relative to the total cross section of the H$(125)$ boson production as predicted in the SM. The observed limit is compatible with the expected limit within one standard deviation in the entire range of $m_{a_1}$ considered, and ranges from 0.022 at $m_{a_1} = 9$ GeV to 0.23 at $m_{a_1} = 4$ GeV. The expected upper limit ranges from 0.027 at $m_{a_1} = 9$ GeV to 0.19 at $m_{a_1} = 15$ GeV. The dependence of the exclusion limit on $m_{a_1}$ has a minimum at 9 GeV. The degradation of the analysis sensitivity towards lower values of $m_{a_1}$ is caused by the increase of the background yield at low invariant masses of the muon-track systems, as illustrated in Figs. 5 and 9. With increasing $m_{a_1}$, the average angular separation between the decay products of the $a_1$ boson is increasing. As a consequence, the efficiency of the signal selection drops down, as we require the muon and the track, originating from the $a_1 \rightarrow \tau \tau$ one-prong decay, to be within a cone of $\Delta R = 0.5$. This explains the deterioration of the search sensitivity at higher values of $m_{a_1}$. The shaded area indicates the excluded value of the branching fraction of the H$(125)$ decay into non-SM particles at 95% CL from Ref. [24]. The new results improve significantly the previous 8 TeV results [26] from 30% (for low masses) up to 80% (for intermediate masses of 8 GeV), while also the new analysis further extends the coverage up to 15 GeV for the considered $a_1$ mass, which were not considered for the previous result.

![Figure 10: The observed and expected upper 95% confidence level limits on the signal cross section times the branching fraction $\sigma(pp \rightarrow H(125) + X) \cdot B(H(125) \rightarrow a_1 a_1) \cdot B^2(a_1 \rightarrow \tau \tau)$, relative to the inclusive Higgs boson production cross section $\sigma_{SM}$ predicted in the SM. The green and yellow bands indicate the regions that contain 68% and 95% of the distribution of upper limits expected assuming no signal is present. The shaded area indicates the excluded value of the branching fraction of the H$(125)$ decay into non-SM particles at 95% CL from Ref. [24].](image)

9 Conclusion

A search for a light pseudoscalar boson $a_1$, produced in decays of the H$(125)$ boson is presented, using a data set corresponding to an integrated luminosity of 35.9 fb$^{-1}$ of proton-proton collisions at a centre-of-mass energy of 13 TeV. The analysis targets the inclusive production of the H$(125)$ boson and exploits the $H(125) \rightarrow a_1 a_1 \rightarrow 4\tau$ decay mode. The contribution of the
H(125) → a_1a_1 → 2μ2τ decay channel is also taken into account in the search. No evidence of signal is found in data. The observed 95% confidence level upper limit on the signal cross section times branching fraction, relative to the cross section of the H(125) boson production predicted in the SM ranges from 0.023 at m_{a_1} = 9 GeV to 0.26 at m_{a_1} = 4 GeV. The expected upper limit ranges from 0.028 at m_{a_1} = 9 GeV to 0.19 at m_{a_1} = 15 GeV.
References


