

# Search for an exotic decay of the 125 GeV Higgs boson to a pair of light pseudoscalars in the final state of two muons and two c-quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV with CMS Open Data

Danyer Perez Adan 

*I. Physikalisches Institut B, RWTH Aachen University,  
Sommerfeldstraße 16, 52074 Aachen, Germany*

*E-mail:* [danyer.perez.adan@rwth-aachen.de](mailto:danyer.perez.adan@rwth-aachen.de)

**ABSTRACT:** A search is performed for pairs of light pseudoscalar bosons (a) produced from decays of the 125 GeV Higgs boson ( $h_{125}$ ). The analysis is based on publicly available data collected in 2016 by the CMS experiment at the LHC in proton-proton collisions at a center-of-mass energy of 13 TeV. The amount of data analyzed corresponds to an integrated luminosity of  $16.4 \text{ fb}^{-1}$ . The analysis explores for the first time at the LHC the final state exhibiting two muons and two c-quarks, which originate from flavor-asymmetric decays of the pseudoscalar pair. The search probes the pseudoscalar boson mass interval comprised between 4 and 11 GeV, which represents a region where the light bosons exhibit a considerable Lorentz boost, and thus their decay products overlap. No significant deviation from the standard model expectation is observed. Model-independent upper limits at 95% confidence level are set on the product of the cross section and branching fraction for the  $h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$  process relative to the standard model Higgs boson production cross section, reaching a minimum value close to  $3.3 \times 10^{-4}$ . The results are interpreted in the context of two Higgs doublets plus singlet models and compared to existing experimental results covering other decay channels. The exclusion limits obtained by this search improve the current constraints set by various LHC searches in scenarios where the coupling of the light boson to up-type quarks is enhanced.

**KEYWORDS:** Hadron-Hadron Scattering , Beyond Standard Model, Higgs Physics

**ARXIV EPRINT:** [2504.12161](https://arxiv.org/abs/2504.12161)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The CMS detector</b>	<b>4</b>
<b>3</b>	<b>Selected and simulated samples</b>	<b>5</b>
<b>4</b>	<b>Event reconstruction and selection</b>	<b>7</b>
<b>5</b>	<b>Signal modeling</b>	<b>15</b>
<b>6</b>	<b>Background modeling</b>	<b>16</b>
<b>7</b>	<b>Systematic uncertainties</b>	<b>18</b>
<b>8</b>	<b>Results</b>	<b>20</b>
<b>9</b>	<b>Summary</b>	<b>24</b>

---

## 1 Introduction

The discovery of a particle [1, 2] exhibiting properties similar to the Higgs boson predicted within the context of the Standard Model (SM) marked the emergence of an entirely new unexplored sector for particle physics. More than twelve years after that remarkable breakthrough, an important number of advances have been made to better understand the nature of such a particle. The mass of the predicted boson has already been measured at a remarkable precision, obtaining a value consistent with 125 GeV and an uncertainty below the 0.1% level [3, 4]. After having been able to observe independently three of the Higgs bosonic decays using the Run 1 data with a statistical significance close to five standard deviations [5–7], later with larger center-of-mass energy and much more collected data during the Run 2 data-taking period, observing and accessing the direct coupling of the Higgs-like particle to the third-generation fermions became a reality [8, 9]. Even on the relatively tiny and experimentally challenging to determine natural width of the Higgs boson, non-negligible constraints have been set already [10, 11]. Experimental studies on the spin and the parity of this new particle have shown compatibility with the SM prediction at a spectacular confidence level [12, 13]. At the current moment, much more refined differential measurements on the various production and decay channels are also available [9, 14], and none of them have evidenced a significant deviation from SM expectations.

Despite all the prominent experimental achievements mentioned above, and the success that it all represents for the SM model, now as a complete theory in its range of validity, it is notorious that the SM alone can not describe many of the experimental observations. Among those pieces of evidence, just to mention a few, there is the presence of an invisible matter (known as *dark matter*) that does not interact electromagnetically, or the matter-antimatter

asymmetry puzzle, which along with some theoretically unsatisfactory aspects of the model that include the absence of the gravitational interaction in its formulation, point to a beyond-SM (BSM) theory. In an attempt to address some of those unanswered questions, countless new models that partly modify the SM structure have been proposed as a potential alternative. With experimental scrutiny of the SM scalar sector having commenced not long ago, it is not a coincidence that the Higgs sector gets particularly influenced in some of the BSM models. This interest is also driven by the intrinsic versatility of scalar fields in incorporating new interactions. Some models propose that the Higgs sector could provide a portal to dark matter [15–17], or may help to generate electroweak baryogenesis of sufficient amount to explain the baryon asymmetry [18, 19]. In the majority of the scenarios, the Higgs sector ends up being augmented, even in minimal extensions of the SM such as supersymmetric models [20] or simple multi-scalar extensions [21]. Requiring an additional SU(2) doublet (see e.g. [22]) is a relatively simple alternative explored in some models, particularly in supersymmetry, due to the ability of this construction to provide mass simultaneously to differently-charged quarks and to cancel anomalies. In general, the structure of these two-Higgs-doublet models (2HDM) can be conceived beyond the particular case of the minimal supersymmetric model (MSSM), giving rise to richer configurations of scalar-to-fermion couplings [23].

Although 2HDMs have received significant constraints by experimental data [24], an extension of these models by additional scalar singlet (2HDM+S) can comfortably circumvent those restrictions if the lightest scalar mass eigenstate is identified as the SM-like 125 GeV state ( $h_{125}$ ) and the model is assumed in the so-called decoupling limit [25]. A concrete realization of such a scalar sector can be found within the context of the next-to-minimal supersymmetric model (NMSSM) [26]. Within this 2HDM+S structure, there are two pseudoscalar states ( $A$  and  $a$ ), one of which (the lightest pseudoscalar, and denoted by  $a$ ) could be very light, even lighter than the SM-like Higgs. Under these assumptions, and if the mass of the lightest pseudoscalar ( $m_a$ ) satisfies the condition  $m_a < m_{h_{125}}/2$ , there could exist exotic decays of the SM-like Higgs of the form  $h_{125} \rightarrow aa$ , where the subsequent decay of the light boson to SM fermions takes place. This decay channel becomes relevant if the lightest pseudoscalar is weakly coupled to other particles (e.g. if  $a$  is mostly-singlet-like), in which cases the primary production of such light pseudoscalars is through Higgs exotic decays. From the experimental standpoint, the current upper bounds at 95% confidence level (CL) on the branching fraction of the Higgs boson to undetected particles set by the ATLAS and CMS experiments are 12% and 16% respectively [8, 9], which still allows for a sufficiently large margin for those exotic decays to exist.

Numerous searches have been performed in the past to look for those exotic Higgs decays. Before the discovery of the SM-like Higgs boson, the D0 collaboration had already looked for  $H \rightarrow aa$  decays in the final states containing muons and tau leptons [27]. Making use of the collected Run 1 data, the CMS and ATLAS collaborations carried out searches in various mass regions [28–32], ranging from very light pseudoscalars (boosted topology) to larger masses (resolved topology) close to half the mass of the by then already found  $h_{125}$  boson. At this point, the number of explored decay channels had already diversified significantly, and experimentally challenging final states such as  $\tau\tau\tau\tau$  and  $\gamma\gamma\gamma\gamma$  were being probed, along with other combinations like  $\mu\mu bb$ , that were added on top of  $\mu\mu\tau\tau$  and  $\mu\mu\mu\mu$  final states. The need to look for different combinations in various decay modes lies in the fact that the exact

configuration of the couplings of the neutral scalar and pseudoscalar states to fermions in the general 2HDM+S can vary and it is unknown a priori [23, 33]. Four types (I, II, III, IV) are typically identified when requiring no flavor-changing neutral currents [23], and within each type, the structure of the coupling to up-type quarks, down-type quarks, and charged leptons is different and dependent on the parameter  $\tan\beta$ .<sup>1</sup>

Considering the above-mentioned complexity, during the Run 2 data-taking period, many more channels were incorporated in the  $h_{125} \rightarrow aa$  search program of the ATLAS and CMS collaborations [31, 34–47]. Final states such as  $bbbb$ ,  $bb\tau\tau$ , and  $\gamma\gamma gg$  were also investigated in diverse production modes for the  $h_{125}$  boson, depending on the experimental requirements for each of the decay channels. At the current moment, despite all the intense effort deployed by the experimental collaborations, no sign of such Higgs exotic decays has been found at a significant level.

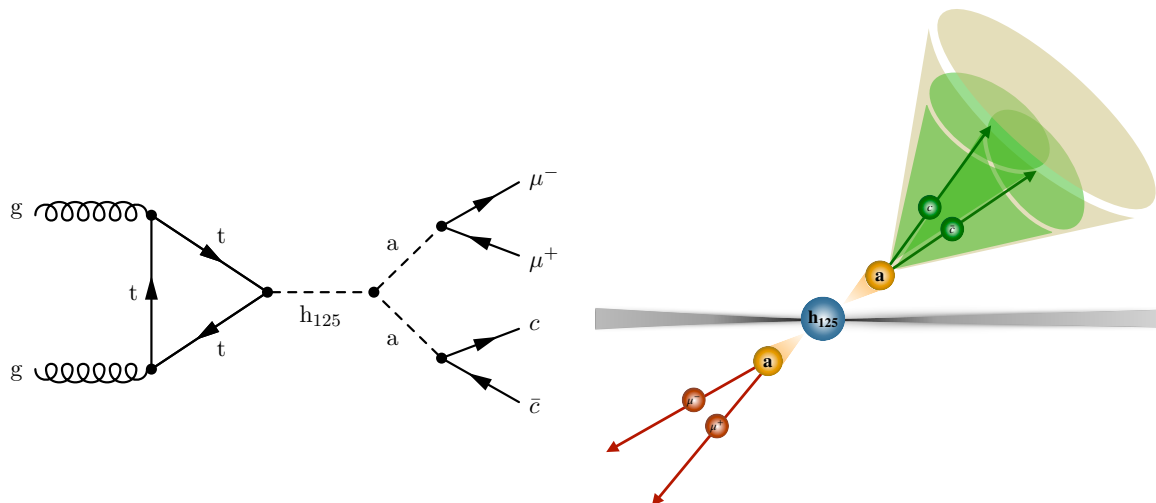
However, despite the multiple searches, there are still regions of the 2HDM+S phase space that none of the already probed decay channels can access. This can be evidenced in some publications where the various ATLAS and CMS analyses are put together [48–50] and projected into specific 2HDM+S configurations. Although the existing searches can cover ample regions in most of the model types, in some cases, like for the Type II and Type III models, none of the channels above can fully access regions of small values of  $\tan\beta$  and very small  $m_a$  (e.g.  $m_a < 12$  GeV). This happens because, in this specific configuration, the decay of the light pseudoscalar is predominantly into c-quarks, light-quarks, and gluons [23]. In particular, for the mass region approximately (with some deviation due to non-perturbative effects [51]) between two times the mass of the c-quark and two times the mass of the b-quark, the decay  $a \rightarrow c\bar{c}$  tends to amply dominate for such low values of the  $\tan\beta$  parameter. Although the use of taggers based on multivariate techniques has not been uncommon [52–54] in the context of  $h_{125} \rightarrow aa$  searches, dedicated developments for hadronic decay channels in more boosted configurations may help make the search with those more involved final states feasible.

The present study performs a dedicated search in the  $h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$  ( $\mu\mu cc$ ) final state using modern charm jet identification techniques [55, 56], which regardless of not being optimal for  $a \rightarrow c\bar{c}$  identification, may still be able to provide enough separation power to explore the uncovered phase space regions. The analysis is based on data collected in 2016 by the CMS experiment in proton-proton (pp) collisions at a center-of-mass energy of 13 TeV, and that was made publicly available by the CMS collaboration [57, 58]. The analysis has been optimized to target the main production mechanism of the SM-like Higgs boson, i.e. gluon fusion (ggF), shown in figure 1 (left), but the contribution arising from vector boson fusion (VBF) mode is also included. The search exploits the invariant mass of the reconstructed  $a \rightarrow \mu^- \mu^+$  candidate to scan for a possible excess over the expected SM background. Masses of the pseudoscalar boson between 4 and 11 GeV are probed, which represents a region where the light boson decay products are collimated and may therefore overlap, in particular, for the  $a \rightarrow c\bar{c}$  leg, as illustrated in figure 1 (right).

This work is structured as follows. A brief description of the CMS detector is provided in section 2. Some details about the simulated MC samples and the chosen datasets are

---

<sup>1</sup>Defined as the ratio of the vacuum expectation value of the second doublet to that of the first doublet.



**Figure 1.** Feynman diagram exemplifying the production and exotic decay of the SM-like Higgs boson into a pair of pseudoscalars that subsequently decay into a  $\mu^- \mu^+$  and  $c \bar{c}$  pair respectively (left). Besides it, there is a schematic representation of the final state topology (right), where the effects of the boosting acquired by the pair of light bosons are illustrated.

discussed in section 3. Section 4 contains a thorough explanation of the event selection employed for this analysis, as well as some essential aspects of the event reconstruction within CMS. The modeling of the di-muon invariant mass for the signal processes here studied is explained in section 5, while in section 6, the model devised for the description of background contributions is covered. Later, in section 7, the treatment of the different sources of systematic uncertainties is presented. Section 8 comprises the various results obtained, and finally, the work presented is summarized in section 9.

## 2 The CMS detector

The data used in this analysis have been recorded with the CMS detector at the LHC in the year 2016. The distinctive component of the CMS detector [59] is a superconducting solenoid of 6 m internal diameter, which is able to supply a magnetic field of 3.8 T. The CMS detector has a cylindrical structure, symmetric around the beam pipe, and centered at the interaction point. The innermost layer is a silicon-based tracker surrounded by a scintillating crystal electromagnetic calorimeter (ECAL). After the ECAL, there is a hadron calorimeter (HCAL) followed by the outermost layer, consisting of systems designed for the detection of muons. More detailed descriptions of the CMS detector, together with a definition of the coordinate system used can be found in ref. [59].

Events of interest are selected using a two-tiered trigger system. The trigger system is responsible for selecting the small fraction of collision events that are relevant to the various physics activities at CMS. The CMS trigger system [59] consists of two stages: the Level-1 trigger [60], which is entirely hardware-based and uses information from the calorimeters and muon detectors to filter events to an output rate of around 100 kHz [61],

and the software-based high-level trigger (HLT) [62] that reduces the rate further down to around 1 kHz before data storage [63].

### 3 Selected and simulated samples

The CMS experiment has released research-quality data since 2014. The CMS Open Data has already been used in several jet substructure studies and BSM searches, as documented in ref. [57]. At the present moment, almost all reconstructed data from Run 1 is already public, along with simulated samples [58]. The first collision data at a center-of-mass energy of 13 TeV was released in 2021, while later, in 2024, about 50% of pp data collected in 2016 at the same center-of-mass energy was released [58]. For detailed information regarding some further aspects of the CMS data preservation and open access policy see refs. [57, 64].

This analysis is based on  $pp$  collisions at a center-of-mass energy of 13 TeV collected by the CMS detector in 2016. The amount of data analyzed is equivalent to a total integrated luminosity of  $16.4 \text{ fb}^{-1}$ . The primary datasets employed contain events recorded with muon triggers, as detailed in refs. [65, 66], and correspond to the last two data-taking eras (labeled  $G$  and  $H$ ) in which the CMS detector collected  $pp$  collisions in 2016.

Although the background estimation method employed in this search is fully data-driven, simulated SM background processes that contribute to the event selection were utilized to perform optimization studies and assess the overall background composition in the various analysis regions defined. As will be detailed in the next section, the most important background sources were found to be quantum chromodynamics production of multi-jet (QCD multi-jet) and Drell-Yan (DY) events. Other minor backgrounds such as top pair production ( $t\bar{t}$ ), single-top associated production with a W boson ( $tW$ ), and di-boson (VV, with  $V = W, Z$ ) production were also included as part of the simulated SM background. Monte Carlo datasets simulated by the CMS collaboration were used for the above-mentioned processes — the full list of samples can be found in table 1. The samples corresponding to VV and QCD were simulated at leading-order (LO) accuracy in QCD by CMS with the PYTHIA [67] (v.8.240) generator using the CP5 tune [68]. For the QCD case, the process was produced differentially in ranges of  $\hat{p}_T$ ,<sup>2</sup> as defined in PYTHIA, with an additional filter at generator level to filter events containing muons with  $p_T > 5 \text{ GeV}$ . The DY samples are generated at LO prediction differentially in di-lepton invariant mass and boson  $p_T$ . For an invariant mass above 10 GeV, the DY samples were generated with up to four partons in the final state using MADGRAPH\_AMC@NLO [69] (v2.6.5) with the MLM prescription [70] for matching jets from the matrix element (ME) calculation to the parton shower description. For the low-mass range in DY production, the same MC generator was used with one parton in the final state, and additionally, the phase space was divided into bins of boson  $p_T$ . Simulated events of  $t\bar{t}$  production and  $tW$  process were generated at next-to-leading order (NLO) in QCD using the POWHEG [71, 72] (v2) event generator. When feasible, the predicted inclusive cross sections of the above simulated processes are corrected to match the most accurate calculations available [73–76], which corresponds to NLO or NNLO predictions for the majority of cases. For all the above samples, PYTHIA was used to simulate parton

<sup>2</sup>The transverse momentum of each of the products, as determined in the rest frame of the colliding partons [67].

Process	Dataset name	Cross section [pb]	Filter eff.
QCD multi-jet	QCD_Pt-15To20_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [79]	$8.488 \times 10^8$	0.003
	QCD_Pt-20To30_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [80]	$3.977 \times 10^8$	0.006
	QCD_Pt-30To50_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [81]	$1.069 \times 10^8$	0.013
	QCD_Pt-50To80_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [82]	$1.572 \times 10^7$	0.024
	QCD_Pt-80To120_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [83]	$2.343 \times 10^6$	0.038
	QCD_Pt-120To170_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [84]	$4.078 \times 10^5$	0.052
	QCD_Pt-170To300_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [85]	$1.036 \times 10^5$	0.068
	QCD_Pt-300To470_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [86]	$6.840 \times 10^3$	0.091
	QCD_Pt-470To600_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [87]	$5.526 \times 10^2$	0.101
	QCD_Pt-600To800_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [88]	$1.566 \times 10^2$	0.116
	QCD_Pt-800To1000_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [89]	$2.628 \times 10^1$	0.125
	QCD_Pt-1000_MuEnrichedPt5_TuneCP5_13TeV-pythia8 [90]	8.230	0.131
DY	DY1jToLL_M-1to10_Pt-0to70_TuneCP5_13TeV-madgraph-pythia8 [91]	$1.279 \times 10^6$	–
	DY1jToLL_M-1to10_Pt-70to100_TuneCP5_13TeV-madgraph-pythia8 [92]	$1.345 \times 10^1$	–
	DY1jToLL_M-1to10_Pt-100to200_TuneCP5_13TeV-madgraph-pythia8 [93]	4.803	–
	DY1jToLL_M-1to10_Pt-200to400_TuneCP5_13TeV-madgraph-pythia8 [94]	0.332	–
	DY1jToLL_M-1to10_Pt-400to600_TuneCP5_13TeV-madgraph-pythia8 [95]	0.014	–
	DY1jToLL_M-1to10_Pt-600toInf_TuneCP5_13TeV-madgraph-pythia8 [96]	0.002	–
		DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8 [97]	$1.861 \times 10^4$
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 [98]	$6.077 \times 10^3$	–
$t\bar{t}$	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 [99]	$8.731 \times 10^1$	–
$tW$	ST_tW_Dilept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8 [100]	7.815	–
VV	WW_TuneCP5_13TeV-pythia8 [101]	$1.187 \times 10^2$	–
	WZ_TuneCP5_13TeV-pythia8 [102]	$4.713 \times 10^1$	–
	ZZ_TuneCP5_13TeV-pythia8 [103]	$1.652 \times 10^1$	–

**Table 1.** List of selected datasets used for SM background processes along with their respective cross section before any event filtering. For those samples in which an event filter at the generator level was utilized, the efficiency values of such requirements are reported in the last column. All above MC samples are provided by the CMS collaboration under ref. [58].

shower, hadronization, and the underlying event [67]. Equally, for all simulated processes, the initial-state partons were modeled with the NNPDF 3.1 NNLO [77] parton distribution function (PDF), while the full CMS detector simulation was performed using GEANT4 [78].

Signal MC samples<sup>3</sup> for the two (ggF and VBF) main production processes of  $h_{125}$ , and where the SM-like Higgs decays via  $h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ , were generated using MADGRAPH\_AMC@NLO [69] (v2.6.5) and the UFO model NMSSMHET provided in ref. [23]. MC samples were produced for several mass hypotheses in the range from 4 to 11 GeV with a step of 1 GeV. To account for a more accurate modeling of the  $p_T$  spectrum of the  $h_{125}$ , the distribution obtained from the MADGRAPH\_AMC@NLO simulation is reweighted to match higher-order predictions. For the ggF process, the HQT program [104] is employed to

<sup>3</sup>These are not part of the CMS Open Data and were therefore generated separately using publicly available information.

compute the  $p_T$  spectrum at NNLL+NLO accuracy, while for the VBF process, the POWHEG generator is used to derive the respective  $p_T$  distribution at NLO precision. The CMS detector geometry and conditions employed in the simulation of the signal processes were taken to be identical to those of the above-described SM backgrounds. The parton shower, hadronization, and the underlying event were simulated using PYTHIA with the embedded CP5 tune. Equally for the PDF, the same version that CMS used for the production of the samples described above was utilized. The previously alluded details, including specific CMS software (CMSSW) version, computing environments, detector conditions, and other relevant elements used to generate the signal samples, are supplied by the CMS collaboration via the CERN Open Data portal [58].

The effects of additional pp interactions in the same or adjacent bunch crossings (pileup) are included in all simulation samples provided by CMS and were also added to the simulation of the signal processes. A reweighting procedure is implemented to match the simulated distribution of pileup interactions with the one observed in the 2016 CMS data.

#### 4 Event reconstruction and selection

The complexity and structure of the released data are the same as the one used by the CMS collaboration [105]. The newly released 2016 collision data and simulation are additionally provided in the so-called NanoAOD format, which embeds standard ROOT [106] data structures. Condensed information for particle candidates that were reconstructed using the CMS particle-flow algorithm is stored for further processing. Legacy conditions of event reconstruction and calibration are typically present in the NanoAOD data format of the released data, although some specific elements of the calibration of the numerous physics objects must be observed when analyzing such a complex dataset. A more detailed explanation and dedicated documentation of the usage of the CMS Open Data can be found in ref. [105].

In the following, a summarized description of the physics object reconstruction within CMS is provided. Emphasis is made on the primary physics objects employed in this search: muons and jets. As documented in ref. [105], the CMS NanoAOD data format already contains the final physics objects as obtained from the particle-flow algorithm, which conducts the fundamental reconstruction steps described below. The requirements utilized to select the physics object candidates are specific to every analysis. The details of those selection criteria when identifying muons and jets in this search are also included below for completeness.

The information provided by the different sub-detectors in CMS is gathered and sent to the next step, which proceeds with the reconstruction and identification of all the stable particles that constitute the event. Other composite-like objects such as jets, missing transverse energy, taus, and primary (secondary) vertices are built, identified, or reconstructed from individual elements.

The particle-flow algorithm [107] is the central element to reconstruct and identify individual particles in a given event. Using a combination of the global information arising from the various elements of the CMS detector (charged particle tracks from the tracking detector, energy deposits in the HCAL and ECAL, and reconstructed tracks from the muon chambers), the multiple possible particles (electrons, muons, photons, charged hadrons, or neutral hadrons) are reconstructed and identified. The reconstructed vertex with the largest

value of summed  $p_T^2$  is taken to be the primary interaction vertex. The main objects used in this analysis are muons and jets, which are briefly discussed in the following.

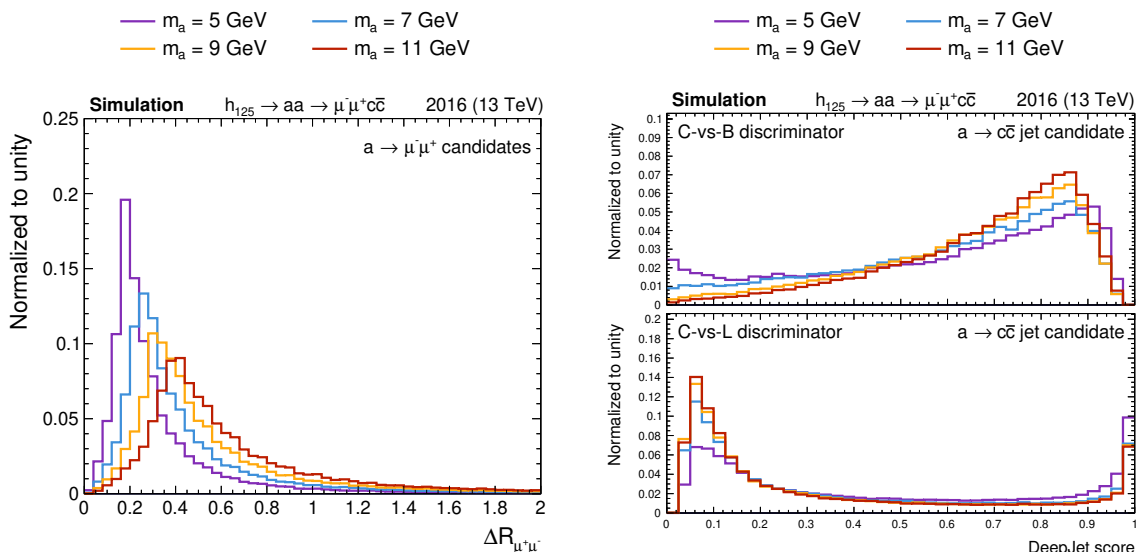
The muons are reconstructed using the information provided by both the tracker and the muon subdetectors, employing a set of dedicated algorithms that identify tracks within the tracker or the muon system, which are later propagated to find potential matches in the alternative subsystem [108]. The muon momentum is obtained from the curvature of the corresponding track by selecting one from several refits to its trajectory based on fit quality and resolution considerations. Within the primary energy range of muons arising from the potential signal here examined, the momentum resolution of muons can be as low as 1% when they are produced in the central part of the detector. In this analysis, muons must pass the “medium” identification criteria [108], designed for high identification efficiency and sufficient background rejection, which corresponds to an approximate 99% efficiency for muons in simulated W and Z events. Muons are required to have  $p_T > 5$  GeV and  $|\eta| < 2.4$ , as well as to pass cuts on the transverse and longitudinal impact parameters of  $d_{xy} < 0.2$  cm and  $d_z < 0.5$  cm respectively. To correct for the difference between simulation and real data, dedicated corrections for muon identification and isolation (see below for an explanation of the usage of isolation in this analysis) efficiencies are applied to simulated events, following the recommendations provided in ref. [105]. These efficiencies have been measured by the CMS collaboration using  $Z \rightarrow \mu^- \mu^+$  (medium-energy muons) and  $J/\psi \rightarrow \mu^- \mu^+$  (low-energy muons) events [108] using the tag-and-probe method. Additionally, and because this analysis relies on the invariant mass reconstruction of a pair of muons, correction factors to improve the calibration of the muon energy scale and resolution are applied.

To reconstruct jets originated by the hadronization of quarks and gluons, firstly an identification of the charged and neutral hadrons that mostly compose them is needed. Charged hadrons are formed by the remaining tracks that do not belong to a muon or electron. Using a matching of the ECAL and HCAL energy deposits together with the track momentum, their energy and momentum can be directly determined. Neutral hadrons are identified by those HCAL energy clusters that are not linked to any charged hadron trajectory, or via a combined ECAL and HCAL energy measurement that exceeds the one expected for a charged hadron energy deposit. The energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy deposits. In this analysis, jets clustered using the anti- $k_T$  clustering algorithm [109] with a distance parameter of 0.4 are used (AK4), employing the corresponding identification and calibration techniques deployed by the CMS collaboration, as explained in ref. [105]. Contamination from pileup and electronic noise is subtracted using the charged-hadron subtraction method [107]. In order to reject jets coming from pileup collisions, a multivariate identification algorithm is applied on relatively low-energetic ( $p_T < 50$  GeV) jets [110]. The energy of reconstructed jets is corrected for effects from the detector response as a function of the jet  $p_T$  and  $\eta$  following the standard procedure [111]. Similarly, to further calibrate the resolution of the energy of reconstructed jets in simulation, a smearing procedure is performed in order to match the observed resolution in real data [111]. In this search, jets must have  $p_T > 25$  GeV and  $|\eta| < 2.4$  to be considered further in the event selection. Jets in the vicinity ( $\Delta R < 0.4$ ) of a selected muon are removed from the analysis, where  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ , with  $\Delta\eta$  and  $\Delta\phi$  the distances in pseudorapidity and azimuthal angle, respectively, between the muon and the jet.

Given the final state studied here, the charm jet identification constitutes a fundamental technique to recognize the  $a \rightarrow c\bar{c}$  decays. Both the identification of  $c$  jets and  $b$  jets in CMS relies on the long lifetime and the mass of the  $c/b$  hadrons, as well as on features of tracks inside the jet (often comprising charged leptons) and the characteristics of reconstructed secondary vertices [56, 112]. Given that the properties of  $c$  jets tend to be somewhere in a middle point between those of light-flavor and  $b$  jets, the identification of  $c$  jets requires the usage of two discriminators. The first one is optimized to distinguish  $c$  jets from light-flavour jets (C-vs-L), whereas the other is trained to distinguish  $c$  jets from  $b$  jets (C-vs-B). In this analysis, the DeepJet algorithm [55] devised within CMS is used as the main instrument to identify  $c$ -jets. This algorithm is a multivariate discriminator based on a deep convolutional neural network architecture. More information on the performance of this classifier in the context of the Run 2 data collected by CMS can be found in ref. [56]. It is common to define several “standard” operating points for those algorithms on which the misidentification probabilities reach a particular value, and that can be used for analysis with different needs in terms of signal purity and efficiency. For the case of the  $c$ -tagging in CMS, three working points (WPs) are defined, based on the bi-dimensional output of the two discriminating variables (C-vs-L and C-vs-B). CMS defines three working points [105] for  $c$ -tagging: loose (L), medium (M), and tight (T). The specific values of both variables that define each WP are documented in ref. [105]. Approximately, in terms of tagging and mistagging rates, the L WP represents an identification efficiency for true  $c$ -jets close to 93%, while it then allows for a misidentification of  $b$  and light jets of 35% and 90% respectively [56]. The T WP, on the other hand, represents an identification efficiency for true  $c$ -jets close to 34%, while it then mistags true  $b$  and light jets with a probability of 20% and 3% respectively [56]. These two previously mentioned WPs are the ones utilized in this search. Corrections to account for the difference in  $c$ -tagging efficiency between simulation and data when operating on these two WPs are applied following the standard procedures and making use of the  $c$ -tagging efficiency measurements laid down by the CMS collaboration [56, 105].

Events are selected using a pair of single-muon triggers, both with a  $p_T$  threshold of 24 GeV, but with slightly different muon object requirements at the online level. The muon objects at the HLT level contain stringent requirements for identification and isolation compared to the generic muon selection described above. In order to match those tighter requirements, the offline muon with the largest transverse momentum (leading muon) is required to match the trigger object. The leading muon is then required to pass the “tight” identification and isolation criteria [108], as well as to have a  $p_T$  larger than 26 GeV. In signal events, the trigger efficiency for muons satisfying such  $p_T$  requirement was found to range from 85% to 92%, depending on the signal mass hypotheses and the transverse momentum of the muons.

At the offline level, events are required to have exactly two oppositely charged muons, one of which, the leading muon, must satisfy the conditions depicted in the above paragraph. The selection continues applying further requirements on the  $p_T$  and  $\Delta R$  of the muons. The figure 2 shows a study performed at generator level, where the characteristics of the  $a \rightarrow \mu^- \mu^+$  and  $a \rightarrow c\bar{c}$  candidates were investigated. On the left side of figure 2, one can observe the distribution of the angular separation ( $\Delta R$ ) between the two muons originating from the  $a \rightarrow \mu^- \mu^+$  leg for a few representative masses of the light boson. One can see that



**Figure 2.** Relevant distributions of studies performed at generator level. On the left, the distribution of the angular separation ( $\Delta R$ ) between the two muons that belong to the  $a \rightarrow \mu^- \mu^+$  decay. On the right, the distribution of the two c-tagging discriminators (C-vs-L and C-vs-B) that are obtained for the DeepJet algorithm applied on the AK4 reconstructed jet that matches the  $a \rightarrow c \bar{c}$  candidate. This matching was performed imposing that the angular distance between the direction of the light boson (truth level) and the direction of the AK4 jet (reconstructed level) is  $\Delta R < 0.4$ . The distributions are normalized to unity and are illustrated for four representative mass points:  $m_a = 5$  GeV (purple line),  $m_a = 7$  GeV (blue line),  $m_a = 9$  GeV (yellow line), and  $m_a = 11$  GeV (red line).

because of the boosting acquired by the pair of pseudoscalars, the two muons become quite collimated — the smaller the mass of the light boson, the smaller the angular separation. This characteristic of the signal was exploited to impose a requirement of  $\Delta R_{\mu^+ \mu^-} < 1$  and  $p_{T, \mu^+ \mu^-} > 40$  GeV on the reconstructed muon pair. From the simulated QCD multi-jet background events, it was possible to verify that, although there is a fraction of events that exhibit a similar low separation between muons, the majority of the events present a softer  $p_{T, \mu^+ \mu^-}$  distribution and a wider  $\Delta R_{\mu^+ \mu^-}$  compared to signal events. Similarly, for the DY background events, even when these tend to have a harder  $p_{T, \mu^+ \mu^-}$  spectrum compared to QCD multi-jet events, the di-muon angular separation in those kinds of events tends to peak for values above 3 (back-to-back configuration).

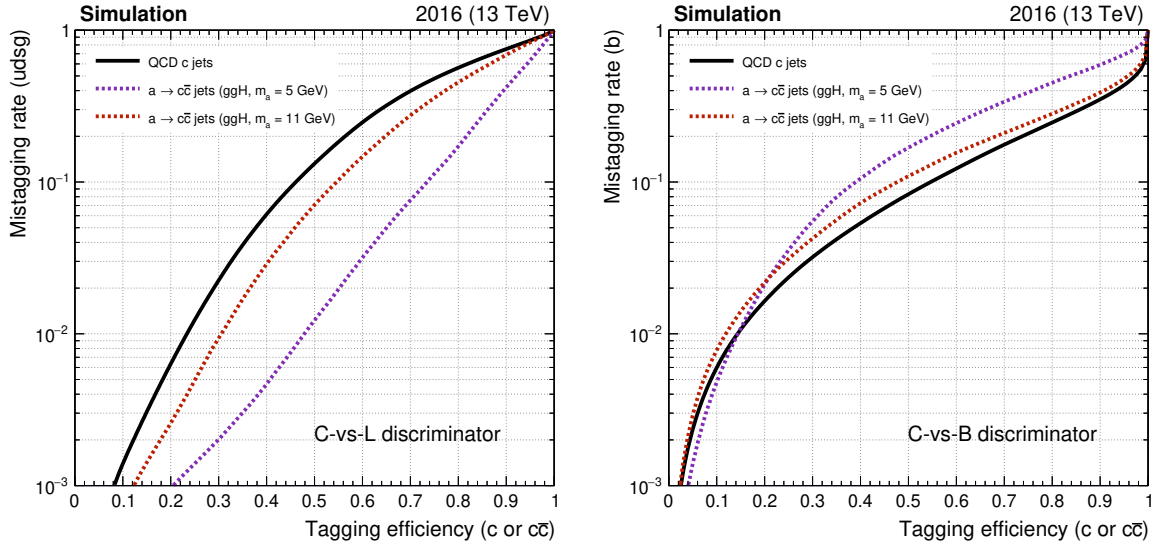
In the next step of the event selection, the analysis employs information about the reconstructed AK4 jets and their c-tagging classification. All the events are required to have at least one reconstructed jet that satisfies the conditions highlighted above for the case of jet objects. Analogously, for the  $a \rightarrow c \bar{c}$ , a dedicated simulation study was performed to evaluate the probability that a single AK4 jet can be reconstructed from the merged decay products of the light boson. The results showed that, in more than 75% of cases, it was possible to reconstruct an AK4 jet that matches in angular distance ( $\Delta R < 0.4$ ) the direction of the pseudoscalar. Then, given this advantageous fact, an evaluation of the output from the classification of the c-tagging DeepJet algorithm was carried out on these particular jets. The findings can be seen in figure 2 (right), where one can note that the distributions of both

the C-vs-L and C-vs-B discriminators when operating on signal  $c\bar{c}$  jets tend to resemble the classification output for true c jets [56]. To quantitatively assess the performance of these two discriminators when applied to signal  $c\bar{c}$  jets, the so-called receiver operating characteristic (ROC) curve is estimated for each of them. The results are shown in figure 3, where the performance of the C-vs-L and C-vs-B discriminators is compared between signal  $c\bar{c}$  jets and regular QCD c jets using the simulated samples. In the case of the signal, only  $c\bar{c}$  jets that are compatible with the direction of the light boson are considered for evaluation. Two substantially different signal mass hypotheses are tested. For the computation of the mistagging rate of true non-c jets, regular QCD light and b jets have been employed. The results in figure 3 indicate that the C-vs-L discriminator successfully identifies signal  $c\bar{c}$  jets with more prominent discrimination against light jets than when operating on regular QCD c jets. This fact is of tremendous advantage for this analysis, given that the most predominant source of background is QCD multi-jet, and naturally, large rates of light jets are normally expected in this background process without any heavy flavor tagging requirement imposed. On the other hand, the C-vs-B discriminator when employed on signal  $c\bar{c}$  jets can barely approach the performance of regular QCD c jets. Nevertheless, this does not prevent the C-vs-B discriminator from effectively distinguishing between signal  $c\bar{c}$  jets and QCD b jets. The above analysis means that there is a non-negligible separation power of the DeepJet c-tagging discriminators for the case of a merged jet formed by decay products coming from the  $a \rightarrow c\bar{c}$  leg. This partial discrimination power of the standard c-tagging algorithm when applied to jets emerging from the  $a \rightarrow c\bar{c}$  leg will be one of the main ingredients used in this analysis to further suppress the background contributions.

In the next stage of the event selection, a requirement of  $p_{T,\text{jet}_L} > 40$  GeV is imposed on the highest- $p_T$  (leading) jet, similarly as also imposed on the reconstructed di-muon pair. This condition allows suppression of both QCD multi-jet and DY background events that typically have a softer  $p_T$  spectrum for jets — the latter was corroborated using the simulated QCD multi-jet and DY samples. Given that the combination of the di-muon pair and the leading jet would reconstruct the  $h_{125}$  candidate, an additional cut was applied on the invariant mass of the object formed by the addition of the four-vector of those three objects. This requirement reads as  $90 \text{ GeV} < m_{\mu^+\mu^-\text{jet}_L} < 160 \text{ GeV}$ , which helps to reduce the QCD multi-jet and DY background by almost a third while keeping the majority of the signal events.

All the above-described event selection, along with an initially loose requirement on the invariant mass of the di-muon pair, namely  $2 \text{ GeV} < m_{\mu^+\mu^-} < 15 \text{ GeV}$ , constitutes the baseline selection for this analysis. A summary of the above-detailed requirements can be found in table 2.

The number of expected and observed events that are obtained with the preliminary selection described above are reported in table 3. The expected background yields are obtained using the reported cross section values in table 1 and the total integrated luminosity recorded by the combination of triggers described above, which amounts to just under  $16.4 \text{ fb}^{-1}$ . For the signal, the reference SM cross sections for the production of the  $h_{125}$  boson in ggF and VBF are used, which are calculated to be 48.58 pb and 3.79 pb respectively [113]. In addition to that, a benchmark branching fraction for the exotic decay of the Higgs to the final state here considered of  $\mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^-\mu^+c\bar{c}) = 10^{-3}$  is assumed, just as a reference for the



**Figure 3.** ROC curves exhibiting the discrimination capability of the C-vs-L (left) and C-vs-B (right) discriminators for the DeepJet algorithm using AK4 jets with similar requirements to the ones utilized in this analysis. The performance of these two discriminators when operating on signal  $c\bar{c}$  jets ( $a \rightarrow c\bar{c}$  jets) is compared to the performance on regular QCD  $c$  jets. For the signal process, only AK4 jets that have been matched by angular distance ( $\Delta R < 0.4$ ) to the direction of flight of the light boson are considered true  $c\bar{c}$  jets. For the estimation of the mistagging rate, regular QCD light ( $udsg$ ) and  $b$  jets have been employed. Curves with dashed lines are shown for two representative mass hypotheses of the signal process:  $m_a = 5$  GeV (purple line) and  $m_a = 11$  GeV (red line). The black solid line has been derived using simulated jets originating from the QCD multi-jet process.

Quantity	Baseline selection
$N_\mu$ (opposite-charge)	= 2
$p_{T,\mu}$	26/5 GeV leading/trailing
$p_{T,\mu^+\mu^-}$	> 40 GeV
$\Delta R_{\mu^+\mu^-}$	< 1
$N_{\text{jet}}$	$\geq 1$
$p_{T,\text{jet}_L}$	> 40 GeV
$m_{\mu^+\mu^-\text{jet}_L}$	$90 \text{ GeV} < m_{\mu^+\mu^-\text{jet}_L} < 160 \text{ GeV}$
$m_{\mu^+\mu^-}$	$2 \text{ GeV} < m_{\mu^+\mu^-} < 15 \text{ GeV}$

**Table 2.** A summary of the main aspects of the baseline event selection for this analysis.

Process	Number of events
QCD multi-jet	$123833 \pm 18600$
DY	$69438 \pm 17626$
$t\bar{t}$	$978 \pm 6$
$tW$	$100 \pm 4$
VV	$72 \pm 3$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 5$ GeV)	$154 \pm 1$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 7$ GeV)	$163 \pm 1$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 9$ GeV)	$169 \pm 1$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 11$ GeV)	$167 \pm 1$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 5$ GeV)	$7.54 \pm 0.07$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 7$ GeV)	$8.05 \pm 0.07$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 9$ GeV)	$8.28 \pm 0.07$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 11$ GeV)	$8.29 \pm 0.07$
Total background	$194421 \pm 25739$
Data	207379

**Table 3.** Expected and observed number of events after the baseline selection in the analysis for the different background and signal processes. The expected background yields are obtained using the reported cross section values in table 1. Typical acceptance (efficiency) values for the different background MC samples reported in table 1 range from  $3 \times 10^{-8}$  to  $2 \times 10^{-5}$  for QCD multi-jet samples, from  $2 \times 10^{-6}$  to  $3 \times 10^{-2}$  for DY samples, from  $7 \times 10^{-4}$  to  $8 \times 10^{-4}$  for  $t\bar{t}$  and  $tW$  samples, and from  $7 \times 10^{-6}$  to  $8 \times 10^{-5}$  for VV samples. For the signal, the SM cross sections for the ggF and VBF processes are used. Additionally, a benchmark branching fraction of  $\mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}) = 10^{-3}$  is assumed. The uncertainties reported are only associated with the statistics of the MC simulation.

values shown in table 1. As one can see, the majority of the background events correspond to QCD multi-jet events, with a contribution close to 70% of the total background. As it will be detailed in section 6, this kind of event corresponds to a large degree to a low-mass resonant QCD background composed of several quarkonium states produced inside jets that can subsequently decay to a pair of muons. The next most important contribution is low-mass DY events with an extra jet, which represents more than 29% of the total background. The contribution from top and di-boson processes is less than 1%.

The final search region where the signal can be extracted is constructed by placing requirements on the c-tagging properties of the leading jet. For these events, the leading jet must pass the T c-tagging WP described before. While this requirement moderately impacts the signal efficiency, reducing it by a multiplicative factor between 0.31 and 0.36 with respect to the baseline selection and depending on the mass hypothesis, the impact

Process	Acceptance ( $\mathcal{A} \times \varepsilon$ )	Number of events
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 5$ GeV)	$0.069 \pm 0.001$	$55.1 \pm 0.6$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 7$ GeV)	$0.061 \pm 0.001$	$48.4 \pm 0.6$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 9$ GeV)	$0.065 \pm 0.001$	$51.7 \pm 0.6$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (ggF, $m_a = 11$ GeV)	$0.067 \pm 0.001$	$53.4 \pm 0.6$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 5$ GeV)	$0.042 \pm 0.001$	$2.61 \pm 0.04$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 7$ GeV)	$0.037 \pm 0.001$	$2.28 \pm 0.04$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 9$ GeV)	$0.038 \pm 0.001$	$2.33 \pm 0.04$
$h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}$ (VBF, $m_a = 11$ GeV)	$0.039 \pm 0.001$	$2.46 \pm 0.04$

**Table 4.** Acceptance (efficiency) values and number of expected events in the SR for a few representative mass hypotheses for the two given  $h_{125}$  production processes. For the calculation of the number of expected events, the SM cross sections for the ggF and VBF processes are used, and a benchmark branching fraction of  $\mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c}) = 10^{-3}$  is assumed. The uncertainties reported are only associated with the statistics of the MC simulation.

on the background is much larger (reduced to a 6% of the number of expected events in the baseline selection), thus allowing to increase the signal over background ratio (S/B) by a factor between 3.9 and 4.5 times with respect to that value after the baseline selection. Finally, and because the probed masses of the light boson range from 4 GeV to 11 GeV, the range of the final discriminant distribution is reduced to  $m_{\mu^+\mu^-} \in [2.5 \text{ GeV}, 12 \text{ GeV}]$ . As it will be discussed in section 6, this choice for the mass range facilitates an adequate coverage of the resonant structure of the QCD multi-jet background, while allowing for an extra margin to fully include signal mass hypotheses close to the edges of the above-defined interval. The region defined by all the above event requirements is henceforth denominated *signal region* (SR). The total number of observed events in the SR is 12996. Unfortunately, due to the insufficient number of simulated events for the main background processes, on top of the intrinsic limitations of the simulation for QCD multi-jet events and low-mass DY events, extracting meaningful values for the expected yields of these processes in the SR turns out to be unfeasible. Relying on the simulation to estimate the full structure and composition of the background for this search is therefore not possible, hence the reason why a completely data-driven method to estimate the background in the SR is devised. For the signal, values of the acceptance and the number of expected events are reported in table 4.

In order to estimate the prevailing shape of the background in the SR, an additional region, where the signal is suppressed with respect to the background, is constructed. This region, denominated as *control region* (CR), is defined by requiring the leading jet to fail the T requirement on the c-tagging classifier, while it is still required to pass the L WP. A simple assessment from simulation confirms that with this condition, the S/B ratio is reduced in the CR by a factor of approximately 10 with respect to the SR. Furthermore, the fact that the differential condition between the SR and the CR is solely based on the  $a \rightarrow c\bar{c}$  jet candidate should be noted. This is of relevance because such a construction makes it

unrelated to the  $a \rightarrow \mu^- \mu^+$  leg. Specifically, the requirement is decorrelated from the  $m_{\mu^+ \mu^-}$  observable, thus allowing to determine the fundamental structure of the background in the SR from this CR — the latter with some caveats, as it will be discussed in section 6. The total number of observed events in the CR is 125709.

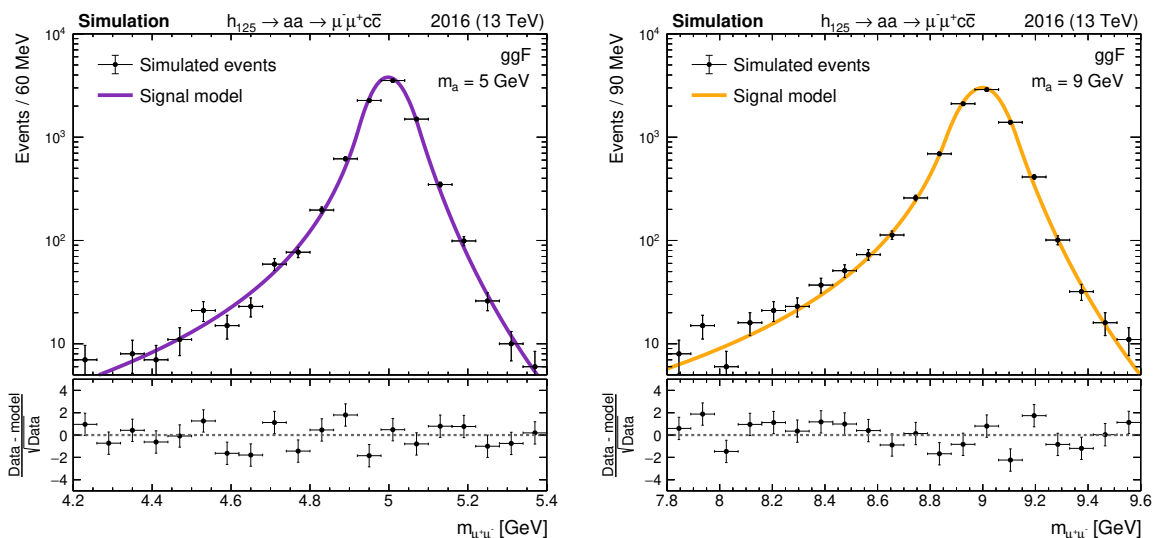
## 5 Signal modeling

As mentioned before, the signal is extracted by fitting the reconstructed di-muon mass distribution. For both the background and the signal, analytic probability density functions (p.d.f.s) are constructed. In order to estimate the functional form of the signal, simulated events are fit to various p.d.f.s that may integrate the different factors affecting the  $m_{\mu^+ \mu^-}$  distribution. It was found that, among the several p.d.f.s examined, the signal shape is well described by a double-sided Crystal Ball function [114], which consists of a double-sided Gaussian central core and a power-law function in each tail portion. The difference with respect to the standard non-double-sided Crystal Ball function is that it contains different parameters for both sides, left (L) and right (R), of the structure. The exact definition can be found in equation (5.1), which is extracted from its implementation in the ROOFIT package [106, 115]. The mathematical formulation of the double-sided Crystal Ball depends on seven parameters and reads

$$f(m_{\mu^+ \mu^-} | m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R, n_L, n_R) = \begin{cases} A_L \cdot \left( B_L - \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_L} \right)^{-n_L}, & \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_L} < -\alpha_L \\ \exp\left(-\frac{1}{2} \cdot \left[ \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_L} \right]^2\right), & \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_L} \leq 0 \\ \exp\left(-\frac{1}{2} \cdot \left[ \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_R} \right]^2\right), & \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_R} \leq \alpha_R \\ A_R \cdot \left( B_R + \frac{m_{\mu^+ \mu^-} - m_0}{\sigma_R} \right)^{-n_R}, & \text{otherwise,} \end{cases} \quad (5.1)$$

with  $A$  and  $B$  normalization parameters, and defined as  $A_{L/R} = \left( \frac{n_{L/R}}{|\alpha_{L/R}|} \right)^{n_{L/R}} \cdot \exp\left(-\frac{|\alpha_{L/R}|^2}{2}\right)$  and  $B_{L/R} = \frac{n_{L/R}}{|\alpha_{L/R}|} - |\alpha_{L/R}|$ . Although, in general, the parametric dependence of equation (5.1) is based on seven parameters, not all of them are necessary to describe the signal shape. In fact, it was found that to reach good modeling of all considered signal scenarios, the most important feature to retain in the above equation was its differential form for the left and right sides. Therefore, it was possible to fix the following four parameters to the values  $\alpha_L = 1.5$ ,  $\alpha_R = 1.5$ ,  $n_L = 2.5$ , and  $n_R = 6.5$ , without affecting the quality of the goodness of fit. An example of the agreement obtained using the above model for two representative mass points and the ggF production mode can be found in figure 4. In general, it was verified that the shape of the signal for a given mass hypothesis does not depend on the  $h_{125}$  production mechanisms here probed, thus the same model was used for both types of processes.

To reproduce the corresponding signal model for intermediate mass points (see section 3) for which a simulated sample was not generated, an interpolation of the three freely floating parameters ( $m_0, \sigma_L, \sigma_R$ ) was performed. It was found that a linear function was sufficient to describe the dependency of these three parameters as a function of the mass hypothesis under consideration. This was validated by generating an intermediate-mass point that was



**Figure 4.** Graphic representation of the unbinned maximum likelihood fit performed to simulated events using the signal model underlined by equation (5.1) and described in the text. Two representative signal mass hypotheses, namely  $m_a = 5$  GeV (left) and  $m_a = 9$  GeV (right), are shown for the ggF process. The data points represent the reconstructed di-muon invariant mass distribution as obtained from simulated signal events, whereas the solid colored line represents the analytical shape of the signal after a fit was performed. The lower panel shows the standard pull obtained when comparing the fitted p.d.f. to the simulated data. The error bars on the data points include only the uncertainty related to the simulation statistics.

not included in the above linear fit and verifying that the predicted model parameters for that point when embedded in equation (5.1) are able to describe the simulated data. On the other hand, for the determination of the acceptance (see table 4) of non-simulated signal mass points, given that the dependency of the signal efficiency in the SR is more complex, a third-degree polynomial was needed. Moreover, since the signal acceptance does depend on the  $h_{125}$  production mode, the determination of the parametric form of the signal acceptance was done differentially for both ggF and VBF processes.

## 6 Background modeling

As explained in section 4, the predominant background sources are QCD multi-jet events and DY low-mass events produced in association with an extra jet. The DY background features a continuum spectrum for the  $m_{\mu^+\mu^-}$  distribution with no relevant resonant structure expected. The QCD multi-jet background, on the other hand, is formed by multiple resonant components and a combinatorial continuum background originated by unrelated sources of opposite charge di-muon candidates. In the mass range studied, there are five prevalent resonances, corresponding to the quarkonium states  $J/\psi(1S)$ ,  $\Psi(2S)$ ,  $\Upsilon(S1)$ ,  $\Upsilon(S2)$ , and  $\Upsilon(S3)$  — with respective approximate masses of 3.096 GeV, 3.686 GeV, 9.460 GeV, 10.023 GeV, and 10.355 GeV, according to [116]. Based on this, the background model is then constructed as the sum of two exponentially decaying functions plus five resonant shapes. The choice for two exponential functions is motivated by the two kinds of continuum backgrounds that

potentially arise from the combinatorial QCD multi-jet background and the low-mass DY events. For the five resonant components, the same double-sided Crystal Ball p.d.f. as for the signal model (see equation (5.1) and related discussion in section 5) is taken. This results in a combined background model depending on 23 different parameters, which can be summarized in the following way:

$$\begin{aligned}
 f(m_{\mu^+\mu^-} | \lambda_1, \lambda_2, \{m_{0,i}, \sigma_{L,i}, \sigma_{R,i}\}, c_i) &= \sum_{i=1}^5 c_i \cdot \mathcal{CB}_i(m_{0,i}, \sigma_{L,i}, \sigma_{R,i}) \\
 &+ c_6 \cdot e^{-\lambda_1 m_{\mu^+\mu^-}} \\
 &+ \left(1 - \sum_{i=1}^6 c_i\right) \cdot e^{-\lambda_2 m_{\mu^+\mu^-}},
 \end{aligned} \tag{6.1}$$

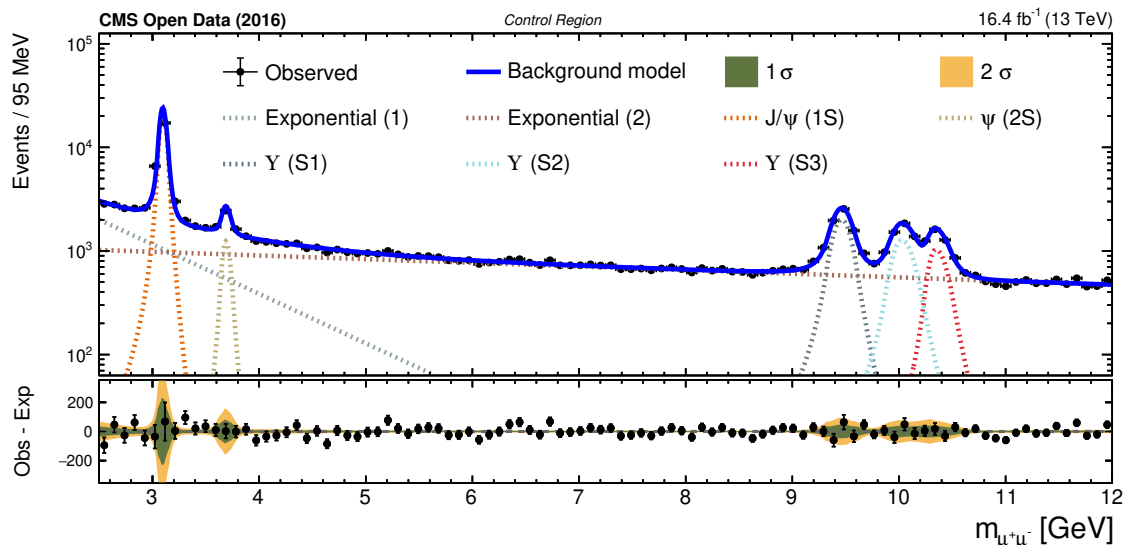
where  $\lambda_1$  and  $\lambda_2$  are the respective exponential decay constants for the two exponential functions,  $c_i$  represent the fraction of each component in the total p.d.f.,<sup>4</sup>  $\mathcal{CB}_i$  corresponds to the Crystal Ball function (equation (5.1)) adopted for each resonances with respective parameter set  $\{m_{0,i}, \sigma_{L,i}, \sigma_{R,i}\}$ , and where  $i = [1, 2, 3, 4, 5] := [J/\psi(1S), \Psi(2S), \Upsilon(1S), \Upsilon(2S), \Upsilon(3S)]$ .

Initially, an unbinned maximum likelihood fit is performed to the CR events, where all the parameters of equation (6.1), as well as the overall background normalization, are left freely floating. The results of this CR-only fit are depicted in figure 5, where a decomposition of the full background model into the different parts that compose it is shown. One can observe that the chosen model can describe the observed data in the CR with a high degree of accuracy, which was corroborated quantitatively ( $p$ -value of 0.12) with a goodness of fit test based on the so-called *saturated model* [117].

Now, the goal is to perform a simultaneous maximum likelihood fit including both the CR and the SR. For this, the background model in the SR is modeled with the same general structure as in the CR, and that is described by equation (6.1), but with some necessary variations. In the SR, where a signal could be present near the quarkonium mass ranges, having all the component fractions associated with the quarkonium contributions freely floating in the fit without any correlation would induce a natural bias in the model, given the similar peak structure for both the signal and the quarkonium states. To mitigate that effect, a correlation between the  $c\bar{c}$  and  $b\bar{b}$  bound states is devised. It is expected that the relation between the number of quarkonium events of a given type that is produced relative to the number of events of another type within the same family would not change between the two regions. Therefore, in the SR, the following restriction for the fractional components are imposed:  $c_{\Psi(2S)}^{\text{SR}} = c_{J/\psi(1S)}^{\text{SR}} \cdot c_{\Psi(2S)}^{\text{CR}} / c_{J/\psi(1S)}^{\text{CR}}$ ,  $c_{\Upsilon(2S)}^{\text{SR}} = c_{\Upsilon(1S)}^{\text{SR}} \cdot c_{\Upsilon(2S)}^{\text{CR}} / c_{\Upsilon(1S)}^{\text{CR}}$ ,  $c_{\Upsilon(3S)}^{\text{SR}} = c_{\Upsilon(1S)}^{\text{SR}} \cdot c_{\Upsilon(3S)}^{\text{CR}} / c_{\Upsilon(1S)}^{\text{CR}}$ . This means that in the combined fit between the SR and the CR the three parameters  $c_{\Psi(2S)}^{\text{SR}}$ ,  $c_{\Upsilon(2S)}^{\text{SR}}$ , and  $c_{\Upsilon(3S)}^{\text{SR}}$  are no longer free, but rather dependent on the remaining quarkonium fractional coefficients in both the SR and CR. That restriction allows for a correlation among the different background resonant components in the SR such that this multi-peak structure can not be biased by a single-peak signal-like appearance.

---

<sup>4</sup>In this case, the p.d.f. is constructed such that it is normalized to unity, thus the fractional coefficient corresponding to the last component is derived from the rest.



**Figure 5.** Illustration of the fit performed to data events selected in the CR using the background model underlined by equation (6.1) and described in the text. The data points represent the reconstructed di-muon invariant mass distribution in the observed data. The blue solid line represents the full background p.d.f. after having performed a fit to CR data. The dotted lines indicate each of the different components that are embedded in the combined background p.d.f.: first exponential function (light gray), second exponential function (brown),  $J/\psi(1S)$  (orange),  $\Psi(2S)$  (olive green),  $\Upsilon(S1)$  (dark gray),  $\Upsilon(S2)$  (cyan), and  $\Upsilon(S3)$  (red). The lower panel shows the difference between the combined background model conceived and the CR data. The green and yellow bands represent a visualization of the 68 and 95% CL uncertainties in the fit. A goodness of fit test based on the *saturated model* revealed a  $p$ -value of 0.12 for the fit depicted in this figure.

The other  $c_i$  parameters, independently in the SR and the CR, are kept freely floating in the combined fit, which includes the fractional components associated with the exponential continuous background — the latter due to the limited accuracy and size of the MC simulation, which made impossible to establish whether the background composition in the CR and SR are the same, thus a more flexible configuration was required. On the other hand, the core shape of the background resonant components is not expected to be different in the CR and SR, therefore, in order to benefit from the higher dataset size in the CR when performing the simultaneous fit, the parameter set  $\{m_{0,i}, \sigma_{L,i}, \sigma_{R,i}\}$  is kept fully correlated (shared) between the two regions. Finally, the overall normalization of the background is inherently expected to change from the CR to SR, so it is naturally kept as an independent and unconstrained parameter in each region during the fit. The full background model described here was tested against potential residual biases using *Asimov* datasets generated by injecting some amount of signal into them — the outcome of that test was successful, and thus no further addition was considered.

## 7 Systematic uncertainties

Since in this analysis the estimation of the background is based on observed data, the modeling of this is not affected by imperfections in the simulation, reconstruction, or detector

response. However, since most of the parameters on which the background p.d.f. model depends, including its normalization, are treated as unconstrained nuisance parameters (see section 6), given the impossibility of imposing restrictions on them based on previous knowledge (e.g. from simulation), they represent the group with the largest impact in the overall sensitivity of the analysis.

On the other hand, several systematic uncertainties affect the modeling of the signal processes here considered. The systematic uncertainties affecting the normalization of the signal are incorporated in the fit via nuisance parameters with a log-normal prior probability density function. The systematic uncertainties that affect the shape of the signal model (see section 5), namely that change the value of the parameters that determine it, are included by adding a direct dependency into the value of the signal p.d.f. parameters, and are assigned a Gaussian prior probability density function.

Multiple uncertainty sources of theoretical origin impact the cross section or the acceptance of the signal processes. In the calculations of the  $h_{125}$  production cross section for ggF and VBF, the unknown contributions from higher-order terms are estimated through variations in the renormalization ( $\mu_R$ ) and factorization ( $\mu_F$ ) scales [113]. This results in a variation of the normalization of the ggF and VBF  $h_{125}$  production modes of up to 6.7% and 0.5% respectively. The impact of the choice of the PDF when performing such calculations was found [113] to change the predicted cross sections, and consequently the normalization of the signal, by 3.2% and 2.1%. The above-described uncertainties comprise only the impact that variations in  $\mu_R$ ,  $\mu_F$ , and PDFs produce in the total cross section used to normalize the signal, however, these variations can also impact the signal acceptance. For the case of the PDF, its impact on signal acceptance was estimated by varying the set in the chosen NNPDF 3.1 NNLO PDF employed in the signal simulation, and it was found to be less than 1% for both processes. In the case of the  $\mu_R$  and  $\mu_F$ , this was done differently for ggF and VBF, though similarly varying both of them within the interval  $0.5 \leq \mu_R/\mu_F \leq 2.0$ . For the ggF, the scales were varied in the HQT program that was used to predict the ggF  $h_{125}$   $p_T$  distribution, and the change that this entailed was propagated to the estimated signal acceptance (see table 4), yielding an overall and non-negligible impact of up to 3.9%. For the VBF process, a similar procedure was followed, but using the respective POWHEG prediction, which in this case represented a change of roughly 1% in the acceptance. All the above theoretical uncertainty sources have no influence on the signal shape.

Among the experimental sources of uncertainties, the determination of the integrated luminosity can vary the total yield of both signal processes up to a value of 2.5% [118]. The 4.8% uncertainty associated with the measurement of the inelastic proton-proton cross section [119] impacts the pileup distribution that is used to reweight the simulated samples, which produces an overall change in the normalization of the signal processes of less than 1%, with negligible effect on the shape of the di-muon invariant mass distribution. Other systematic uncertainties such as the muon identification, isolation, and trigger efficiency associated with the leading muon have a similar order of magnitude ( $< 1\%$ ) in the impact on the signal acceptance. On the other hand, a quite important change of between 15–17% in the acceptance, depending on the particular signal process, was found to arise from the identification of the less energetic muon in the event. Such an effect might be related to the larger uncertainties [108] obtained in the calibration measurements performed by the

CMS collaboration when determining the identification efficiency of low- $p_T$  muons. Another relevant impact connected to muon objects, in this case affecting the shape of the signal, is the one related to the uncertainty of the muon energy scale and resolution. This implies an increase or decrease of the  $p_T$  of muons, which in the end produces a shift in the di-muon invariant mass distribution. The three free parameters  $\{m_0, \sigma_L, \sigma_R\}$  in the signal p.d.f. are affected by this uncertainty, and therefore, a functional dependency for these three parameters on a nuisance parameter assigned to the muon scale/resolution, and determined by refitting the signal for respective variations within the measured scale/resolution uncertainties, was incorporated in the fit. The remaining sources of systematic uncertainties are related to jet objects, and can thus only affect the acceptance of the signal. Among these, and as expected, the most relevant effect arises from the uncertainty on the c-tagging efficiency. The latter produces a total change of 4.7% and 6.7% on the signal acceptance of the ggF and VBF processes respectively. The uncertainty on the efficiency of the jet pileup identification algorithm and the uncertainty related to the jet energy resolution generate minimal effects on the normalization of less than 1%. The uncertainties related to the jet energy scale give rise to a change in the acceptance of 1.2% and 0.5% for the ggF and VBF processes respectively.

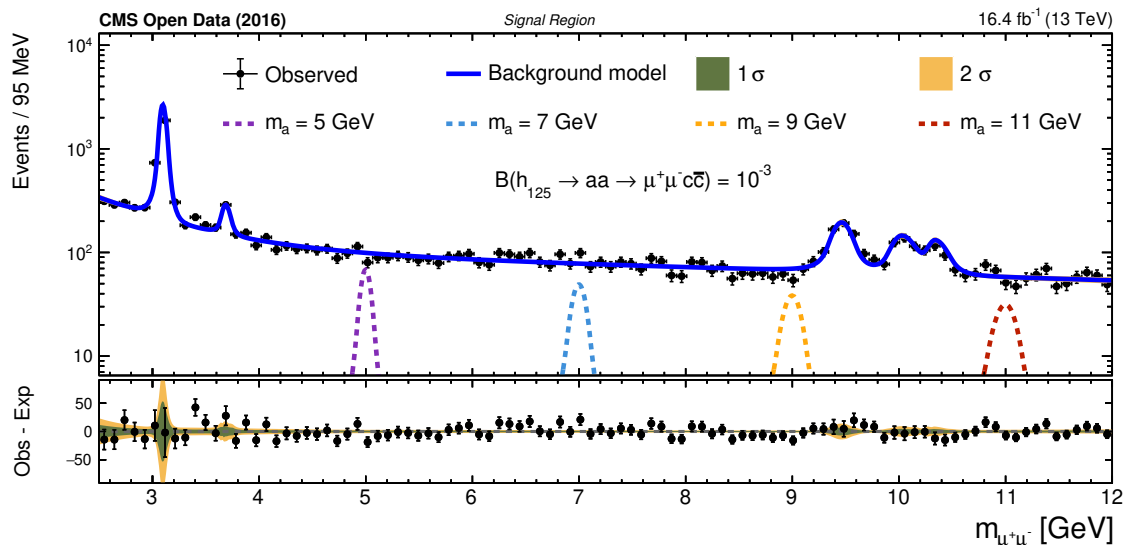
Finally, the uncertainty associated with the limited simulation statistics in the signal samples can marginally impact the values of the model parameters and the acceptance determination. The magnitude of the impact of the simulation statistics on the acceptance can be seen in table 4, while to account for the impact on the signal p.d.f. parameters, the post-fit uncertainties obtained were appropriately incorporated as dependencies of  $\{m_0, \sigma_L, \sigma_R\}$  on three additional nuisance parameters.

## 8 Results

As initially mentioned in section 1, the di-muon invariant mass is scanned in a search for a potential excess of signal events, for which an unbinned maximum-likelihood fit to this distribution is performed using a combination of selected events in both the SR and CR. In this fit, both the signal normalization and the respective background normalizations (see section 6) are left freely floating. All other signal p.d.f. parameters are only allowed to vary within their respective uncertainties for a given mass hypothesis, as described in section 7. For the case of the background, all parameters included in equation (6.1), taking into account the particular region that they represent, are left unconstrained as well. The only constraints incorporated into the p.d.f. model describing background events in the SR are the ones above-mentioned in section 6.

To ensure objectivity in this search, the data in the SR was kept hidden (“blinded”) during the analysis. The optimization of the event selection (see section 4) for the decay channel examined in this work has been based purely on simulation. Furthermore, several diagnostic checks to investigate the robustness of the statistical model, aside from the CR-only fit described in section 6, were performed by generating pseudo-data in the SR [117].

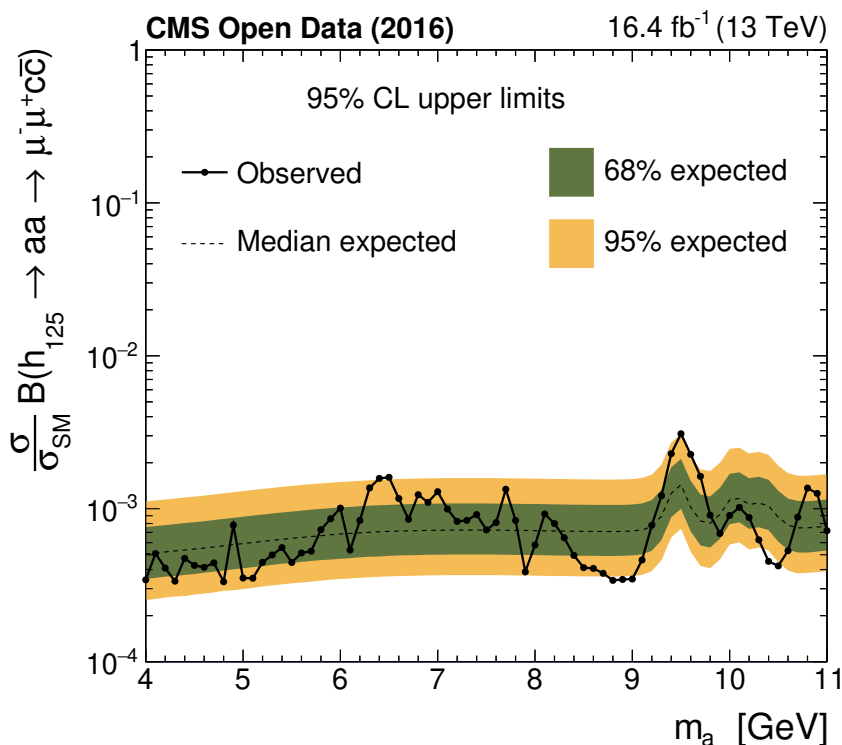
In figure 6, the di-muon invariant mass distribution in the SR is shown. The figure shows the obtained background profile after applying a simultaneous fit to the observed data in both the CR and the SR under the background-only hypothesis, as well as the expectations for the signal for a few representative mass points. As can be seen, no significant deviations from the expected SM background are observed in that distribution.



**Figure 6.** Invariant mass distribution of the muon pair in the SR after having performed a simultaneous background-only fit to the observed events in the SR and the CR. The black points represent the observed data, which has been binned for illustration purposes. The blue solid line represents the background prediction in the SR, while the dashed lines indicate the expected curves for four representative signal mass hypotheses:  $m_a = 5$  GeV (purple),  $m_a = 7$  GeV (cyan),  $m_a = 9$  GeV (yellow), and  $m_a = 11$  GeV (red). For this graphical representation, the signal has been normalized to the sum of the SM cross sections of the ggF and VBF processes, multiplied by a benchmark branching fraction of  $\mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^-\mu^+c\bar{c}) = 10^{-3}$ . The lower panel shows the difference between the observed data in the SR and the expected SM background. The green and yellow bands represent a visualization of the 68 and 95% CL uncertainties in the fit. A goodness of fit test based on the *saturated model* revealed a  $p$ -value of 0.36 for the simultaneous fit illustrated above.

Upper limits at 95% CL are set on the combined product of the production cross section and branching fraction relative to the SM Higgs boson production cross section, namely  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^-\mu^+c\bar{c})$ , for pseudoscalar masses between 4 and 11 GeV. The limits are computed using the modified frequentist  $\text{CL}_s$  approach [120, 121] employing an asymptotic approximation to the distribution of the profile likelihood ratio test statistic [119]. The results are presented in figure 7, and they corroborate the above observation made when performing a maximum-likelihood fit under the background-only hypothesis, namely, that no significant excesses are seen. Only a few minor and local deviations, very close to the two-standard-deviation threshold are preferred by the data when scanning for potential signal hypotheses. As a result, the median expected upper limit ranges from  $5.0 \times 10^{-4}$  for a mass close to  $m_a = 4$  GeV up to a value of  $1.4 \times 10^{-3}$  for a mass hypothesis in the vicinity of  $m_a = 9.5$  GeV, while the observed upper limits ranges from  $3.3 \times 10^{-4}$  for a mass close to  $m_a = 4.8$  GeV up to a value of  $3.1 \times 10^{-3}$  for  $m_a = 9.5$  GeV.

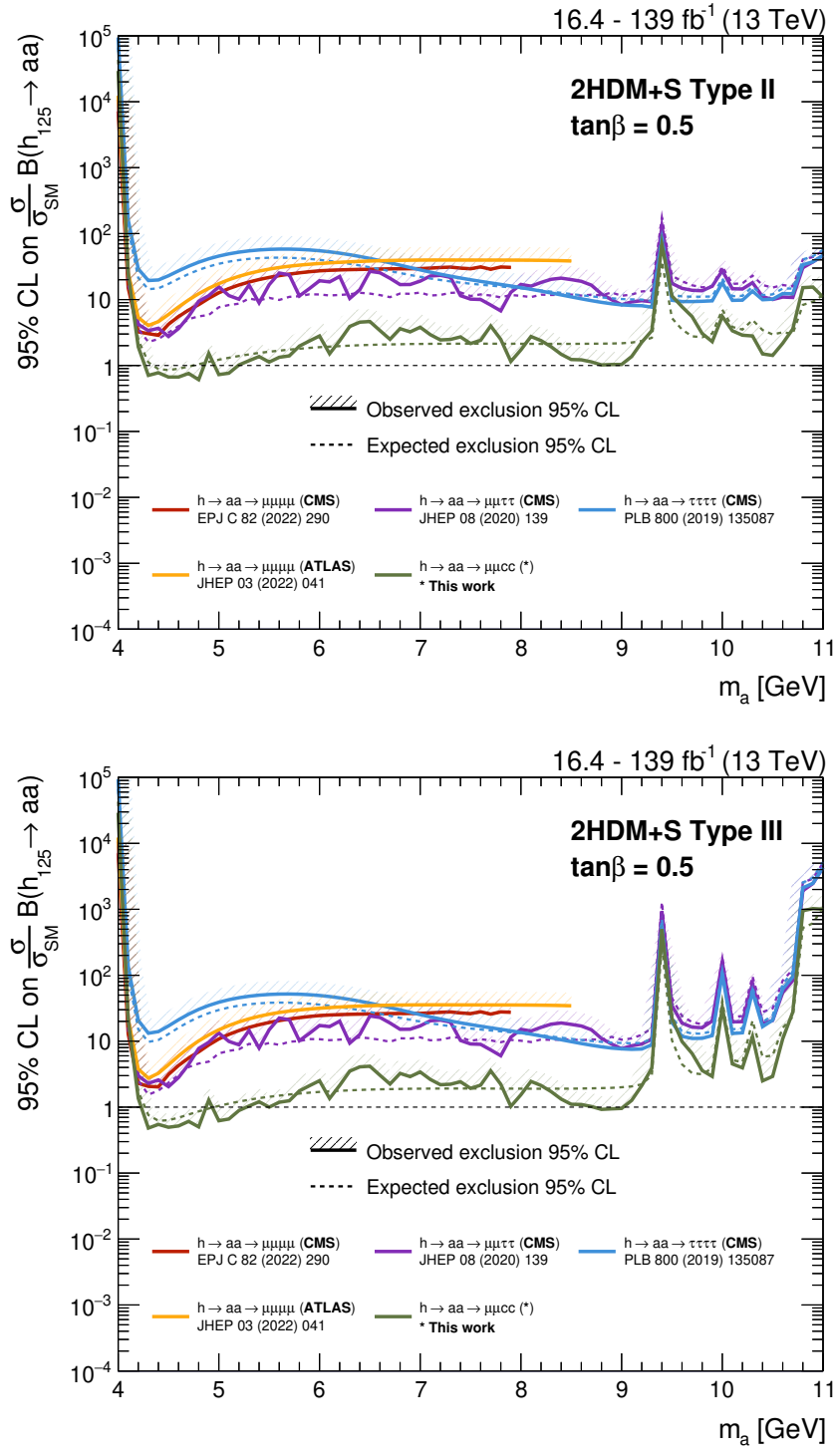
As can be noted, the analysis expected sensitivity slightly degrades as the mass of the light boson increases. This can be explained by the fact that, despite the background is falling roughly exponentially from  $m_a \approx 4$  GeV to  $m_a \approx 9$  GeV, the narrower signal peak featured by lighter states, makes the overall discrimination more prominent for masses in the lower range of the search interval. From  $m_a \approx 9$  GeV to  $m_a \approx 10.5$  GeV, the resonant components of the



**Figure 7.** Observed and expected upper limits at 95% CL on the product of the signal cross section and branching fraction  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^- \mu^+ c\bar{c})$  relative to the SM prediction. The solid and dashed lines correspond to the observed and median expected limits, respectively, while the green and yellow bands indicate the regions that contain 68% and 95% of the expected upper limits.

background predominate over the exponential continuum, which consequently engenders a further deterioration of the expected limits in that mass range.

The above limits presented in figure 7 can be regarded as model-independent under the assumption that the narrow width approximation is valid for all resonances involved in the decay chain — a condition that is comfortably applicable to both  $h_{125}$  and  $a$ . The results are thus translated into model-dependent constraints on  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa)$  as a function of  $m_a$ , and assuming a value of  $\tan \beta = 0.5$ , for Type II and III scenarios of the 2HDM+S. The choice for this particular value of  $\tan \beta$  is motivated by the observations made in section 1, and it exemplifies a point embedded in the phase-space region for which both model types above feature a preferred coupling of the light pseudoscalar to up-type quarks, or simultaneously, to both up-type and down-type quarks. For the reinterpretation of the results shown in figure 7, the model branching fractions for pseudoscalar decays to a pair of fermions,  $\mathcal{B}(a \rightarrow f\bar{f})$ , were taken from [51]. To profit from a higher mass granularity compared to the existing scanned mass points, a spline interpolation has been performed for both the theoretical predictions of  $\mathcal{B}(a \rightarrow f\bar{f})$  and the available experimental limits. The upper limits at 95% CL on  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa)$  can be found in figure 8, along with a comparison with the sensitivity of several experimental results mentioned in section 1, and that are relevant for the mass range probed in this work. Specifically, results from both the CMS and the ATLAS collaborations using at least  $35.9 \text{ fb}^{-1}$  of data that cover the  $\mu\mu\mu\mu$ ,  $\mu\mu\tau\tau$ , and  $\tau\tau\tau\tau$  final states are included in the figure along with the results obtained in this analysis.



**Figure 8.** Observed and expected upper limits at 95% CL on  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa)$  as a function of  $m_a$  for Type II (upper) and Type III (lower) 2HDM+S scenarios. The limits are computed assuming a value of  $\tan\beta = 0.5$ . The results of this search employing the CMS Open Data ( $16.4 \text{ fb}^{-1}$ ) are compared to several experimental results delivered by both the CMS and the ATLAS collaborations in the  $\mu\mu\mu\mu$ ,  $\mu\mu\tau\tau$ , and  $\tau\tau\tau\tau$  final states. The CMS and ATLAS results comprise at least  $35.9 \text{ fb}^{-1}$  of data.

As can be observed in figure 8, for these two 2HDM+S scenarios, the constraints imposed by this analysis are much more stringent than those of the other existing analyses for the same mass range. In fact, it can be said that this analysis offers the first physical constraints for  $\mathcal{B}(h_{125} \rightarrow aa)$ , if a similar cross section to the one predicted in the SM for the Higgs is assumed. The stronger sensitivity of the  $\mu\mu cc$  channel compared to fully leptonic final states can be understood if one considers that for the 2HDM+S configurations shown in figure 8, the light boson predominantly decays to a pair of c-quarks, given its preferred coupling to up-type quarks. Leptonic decays of the pseudoscalar, despite the experimental advantages offered, are highly suppressed in those particular 2HDM+S scenarios. A combination of high decay rates via the  $a \rightarrow c\bar{c}$  mode and the excellent experimental reconstruction achieved when probing  $a \rightarrow \mu^-\mu^+$  decays thus delivers a more powerful constraint in the above-depicted 2HDM+S parameter phase space. Up to the present moment, this search is the only experimental result that is able to cross the unity threshold for  $\mathcal{B}(h_{125} \rightarrow aa)$ , under the specific model configurations explained above. These constraints are still limited to a small mass interval though. However, it is still plausible to think of further improvements for the mass range corresponding to heavier pseudoscalars, as more of the data that has already been recorded in Run 2 and during Run 3 could be added. The development of dedicated c-taggers optimized for the specific topology studied here could also contribute substantially to strengthening the potential of such a challenging but fascinating final state.

## 9 Summary

The first search for exotic decays of the 125 GeV Higgs boson ( $h_{125}$ ) into a pair of light bosons (a) in the  $h_{125} \rightarrow aa \rightarrow \mu^-\mu^+c\bar{c}$  channel has been presented. A publicly available dataset of proton-proton collisions collected in 2016 by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to a total integrated luminosity of  $16.4 \text{ fb}^{-1}$ , was analyzed.

The analysis exploits, for the first time in these types of searches, the prospect of charm jet identification techniques when applied to collimated and low-mass  $a \rightarrow c\bar{c}$  systems. The current c-tagging methods employed by the CMS collaboration, even when not mainly designed to tackle such a class of topologies, can identify a variety of these scenarios with adequate efficiency. The above, when combined with the powerful di-muon mass resolution exhibited by the CMS detector, allows us to reach considerable levels of sensitivity for this kind of process.

No sign of decays of the 125 GeV scalar into a pair of pseudoscalars via the channel investigated here has been observed. The results are thus presented in terms of 95% CL upper limits on the product of the cross section and branching fraction relative to the standard model (SM) Higgs boson production cross section, i.e.  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa \rightarrow \mu^-\mu^+c\bar{c})$ . The above limits are translated into model-specific constraints on  $\sigma/\sigma_{\text{SM}} \cdot \mathcal{B}(h_{125} \rightarrow aa)$  in the context of Type II and III two Higgs doublets plus singlet models (2HDM+S). The exclusion limits established by this search are compared to several experimental results obtained by the ATLAS and CMS collaborations in other decay channels, which make use of sizably larger datasets. By probing those 2HDM+S configurations where the coupling of up-type quarks to the light pseudoscalar is enhanced, it is demonstrated that this search produces the most stringent constraints up to date for those model realizations.

## Acknowledgments

D.P.A. would like to thank his colleagues from the CMS collaboration for releasing about half of the total dataset collected in 2016, as well as a large amount of different simulated processes as part of the Open Data. D.P.A. acknowledges the dedicated effort to document the details of the usage of the software, detector calibration, and other relevant aspects of the released data through the various public web portals. Furthermore, D.P.A. would like to express his gratitude to the CMS collaboration for permitting him to pursue this publication under individual authorship. This work is possible thanks to the outstanding performance of the LHC through these years and the impressive computing infrastructure deployed and supported by CERN. This work has been supported by the Bundesministerium für Bildung und Forschung. Finally, D.P.A. genuinely values the discussions with his colleagues from the I. Physikalisches Institut B at RWTH Aachen and the insightful comments received for the completion of this manuscript.

**Data Availability Statement.** This article has no associated data or the data will not be deposited.

**Code Availability Statement.** This article has no associated code or the code will not be deposited.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [[arXiv:1207.7214](https://arxiv.org/abs/1207.7214)] [[INSPIRE](#)].
- [2] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [[arXiv:1207.7235](https://arxiv.org/abs/1207.7235)] [[INSPIRE](#)].
- [3] CMS collaboration, *Measurement of the Higgs boson mass and width using the four-lepton final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **111** (2025) 092014 [[arXiv:2409.13663](https://arxiv.org/abs/2409.13663)] [[INSPIRE](#)].
- [4] ATLAS collaboration, *Combined Measurement of the Higgs Boson Mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  Decay Channels with the ATLAS Detector Using  $\sqrt{s} = 7, 8,$  and 13 TeV  $pp$  Collision Data*, *Phys. Rev. Lett.* **131** (2023) 251802 [[arXiv:2308.04775](https://arxiv.org/abs/2308.04775)] [[INSPIRE](#)].
- [5] CMS collaboration, *Measurement of Higgs Boson Production and Properties in the  $WW$  Decay Channel with Leptonic Final States*, *JHEP* **01** (2014) 096 [[arXiv:1312.1129](https://arxiv.org/abs/1312.1129)] [[INSPIRE](#)].
- [6] CMS collaboration, *Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State*, *Phys. Rev. D* **89** (2014) 092007 [[arXiv:1312.5353](https://arxiv.org/abs/1312.5353)] [[INSPIRE](#)].
- [7] CMS collaboration, *Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties*, *Eur. Phys. J. C* **74** (2014) 3076 [[arXiv:1407.0558](https://arxiv.org/abs/1407.0558)] [[INSPIRE](#)].

- [8] CMS collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60 [[arXiv:2207.00043](#)] [[INSPIRE](#)].
- [9] ATLAS collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52 [*Erratum ibid.* **612** (2022) E24] [[arXiv:2207.00092](#)] [[INSPIRE](#)].
- [10] CMS collaboration, *Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production*, *Nature Phys.* **18** (2022) 1329 [[arXiv:2202.06923](#)] [[INSPIRE](#)].
- [11] ATLAS collaboration, *Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector*, *Phys. Lett. B* **846** (2023) 138223 [*Erratum ibid.* **854** (2024) 138734] [[arXiv:2304.01532](#)] [[INSPIRE](#)].
- [12] CMS collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, *Phys. Rev. D* **92** (2015) 012004 [[arXiv:1411.3441](#)] [[INSPIRE](#)].
- [13] ATLAS collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 476 [*Erratum ibid.* **76** (2016) 152] [[arXiv:1506.05669](#)] [[INSPIRE](#)].
- [14] ATLAS and CMS collaborations, *Interpretation of ATLAS and CMS Higgs measurements in STXS and EFT*, *PoS LHCP2020* (2021) 137 [[INSPIRE](#)].
- [15] G. Arcadi, A. Djouadi and M. Raidal, *Dark Matter through the Higgs portal*, *Phys. Rept.* **842** (2020) 1 [[arXiv:1903.03616](#)] [[INSPIRE](#)].
- [16] R. Soualah and A. Ahriche, *Scale invariant scotogenic model: Dark matter and the scalar sector*, *Phys. Rev. D* **105** (2022) 055017 [[arXiv:2111.01121](#)] [[INSPIRE](#)].
- [17] E. Cervantes, O. Perez-Figueroa, R. Perez-Martinez and S. Ramos-Sanchez, *Higgs-portal dark matter from nonsupersymmetric strings*, *Phys. Rev. D* **107** (2023) 115007 [[arXiv:2302.08520](#)] [[INSPIRE](#)].
- [18] D.E. Morrissey and M.J. Ramsey-Musolf, *Electroweak baryogenesis*, *New J. Phys.* **14** (2012) 125003 [[arXiv:1206.2942](#)] [[INSPIRE](#)].
- [19] D. Bodeker and W. Buchmuller, *Baryogenesis from the weak scale to the grand unification scale*, *Rev. Mod. Phys.* **93** (2021) 035004 [[arXiv:2009.07294](#)] [[INSPIRE](#)].
- [20] S.P. Martin, *A Supersymmetry primer*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1 [[hep-ph/9709356](#)] [[INSPIRE](#)].
- [21] A. Drozd, B. Grzadkowski and J. Wudka, *Multi-Scalar-Singlet Extension of the Standard Model — the Case for Dark Matter and an Invisible Higgs Boson*, *JHEP* **04** (2012) 006 [*Erratum ibid.* **11** (2014) 130] [[arXiv:1112.2582](#)] [[INSPIRE](#)].
- [22] G.C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1 [[arXiv:1106.0034](#)] [[INSPIRE](#)].
- [23] D. Curtin et al., *Exotic decays of the 125 GeV Higgs boson*, *Phys. Rev. D* **90** (2014) 075004 [[arXiv:1312.4992](#)] [[INSPIRE](#)].
- [24] G. Belanger et al., *Global fit to Higgs signal strengths and couplings and implications for extended Higgs sectors*, *Phys. Rev. D* **88** (2013) 075008 [[arXiv:1306.2941](#)] [[INSPIRE](#)].
- [25] J.F. Gunion and H.E. Haber, *The CP conserving two Higgs doublet model: The approach to the decoupling limit*, *Phys. Rev. D* **67** (2003) 075019 [[hep-ph/0207010](#)] [[INSPIRE](#)].

- [26] U. Ellwanger, C. Hugonie and A.M. Teixeira, *The Next-to-Minimal Supersymmetric Standard Model*, *Phys. Rept.* **496** (2010) 1 [[arXiv:0910.1785](#)] [[INSPIRE](#)].
- [27] D0 collaboration, *Search for NMSSM Higgs bosons in the  $h \rightarrow aa \rightarrow \mu\mu\mu\mu/\mu\mu\tau\tau$  channels using  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Rev. Lett.* **103** (2009) 061801 [[arXiv:0905.3381](#)] [[INSPIRE](#)].
- [28] CMS collaboration, *A search for pair production of new light bosons decaying into muons*, *Phys. Lett. B* **752** (2016) 146 [[arXiv:1506.00424](#)] [[INSPIRE](#)].
- [29] CMS collaboration, *Search for a very light NMSSM Higgs boson produced in decays of the 125 GeV scalar boson and decaying into  $\tau$  leptons in  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, *JHEP* **01** (2016) 079 [[arXiv:1510.06534](#)] [[INSPIRE](#)].
- [30] CMS collaboration, *Search for light bosons in decays of the 125 GeV Higgs boson in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, *JHEP* **10** (2017) 076 [[arXiv:1701.02032](#)] [[INSPIRE](#)].
- [31] ATLAS collaboration, *Search for Higgs bosons decaying to  $aa$  in the  $\mu\mu\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS experiment*, *Phys. Rev. D* **92** (2015) 052002 [[arXiv:1505.01609](#)] [[INSPIRE](#)].
- [32] ATLAS collaboration, *Search for new phenomena in events with at least three photons collected in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 210 [[arXiv:1509.05051](#)] [[INSPIRE](#)].
- [33] N. Craig et al., *Multi-Lepton Signals of Multiple Higgs Bosons*, *JHEP* **02** (2013) 033 [[arXiv:1210.0559](#)] [[INSPIRE](#)].
- [34] CMS collaboration, *Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state of two muons and two  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **11** (2018) 018 [[arXiv:1805.04865](#)] [[INSPIRE](#)].
- [35] CMS collaboration, *Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two  $b$  quarks and two  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Lett. B* **785** (2018) 462 [[arXiv:1805.10191](#)] [[INSPIRE](#)].
- [36] CMS collaboration, *A search for pair production of new light bosons decaying into muons in proton-proton collisions at 13 TeV*, *Phys. Lett. B* **796** (2019) 131 [[arXiv:1812.00380](#)] [[INSPIRE](#)].
- [37] CMS collaboration, *Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two muons and two  $b$  quarks in  $pp$  collisions at 13 TeV*, *Phys. Lett. B* **795** (2019) 398 [[arXiv:1812.06359](#)] [[INSPIRE](#)].
- [38] CMS collaboration, *Search for light pseudoscalar boson pairs produced from decays of the 125 GeV Higgs boson in final states with two muons and two nearby tracks in  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Lett. B* **800** (2020) 135087 [[arXiv:1907.07235](#)] [[INSPIRE](#)].
- [39] CMS collaboration, *Search for a light pseudoscalar Higgs boson in the boosted  $\mu\mu\tau\tau$  final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **08** (2020) 139 [[arXiv:2005.08694](#)] [[INSPIRE](#)].
- [40] CMS collaboration, *Search for the exotic decay of the Higgs boson into two light pseudoscalars with four photons in the final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **07** (2023) 148 [[arXiv:2208.01469](#)] [[INSPIRE](#)].
- [41] CMS collaboration, *Search for the decay of the Higgs boson to a pair of light pseudoscalar bosons in the final state with four bottom quarks in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **06** (2024) 097 [[arXiv:2403.10341](#)] [[INSPIRE](#)].

- [42] CMS collaboration, *Search for low-mass dilepton resonances in Higgs boson decays to four-lepton final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **82** (2022) 290 [[arXiv:2111.01299](#)] [[INSPIRE](#)].
- [43] ATLAS collaboration, *Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with  $139\text{fb}^{-1}$  of pp collision data at  $\sqrt{s} = 13$  TeV*, *JHEP* **03** (2022) 041 [[arXiv:2110.13673](#)] [[INSPIRE](#)].
- [44] ATLAS collaboration, *Search for Higgs boson decays into a pair of pseudoscalar particles in the  $bb\mu\mu$  final state with the ATLAS detector in pp collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **105** (2022) 012006 [[arXiv:2110.00313](#)] [[INSPIRE](#)].
- [45] ATLAS collaboration, *Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the  $H \rightarrow aa \rightarrow 4b$  channel in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **10** (2018) 031 [[arXiv:1806.07355](#)] [[INSPIRE](#)].
- [46] ATLAS collaboration, *Search for Higgs boson decays into two new low-mass spin-0 particles in the  $4b$  channel with the ATLAS detector using pp collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **102** (2020) 112006 [[arXiv:2005.12236](#)] [[INSPIRE](#)].
- [47] ATLAS collaboration, *Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the  $\gamma\gamma jj$  final state in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **782** (2018) 750 [[arXiv:1803.11145](#)] [[INSPIRE](#)].
- [48] ATLAS collaboration, *Summary of Exotic Higgs Boson Decays from the ATLAS Experiment*, [ATL-PHYS-PUB-2025-011](#), CERN, Geneva (2025).
- [49] ATLAS et al. collaborations, *Highlights on Supersymmetry and Exotic Searches at the LHC*, in the proceedings of the *32nd Rencontres de Blois on Particle Physics and Cosmology*, Blois, France, October 17–22 (2021) [[arXiv:2204.03053](#)] [[INSPIRE](#)].
- [50] ATLAS and CMS collaborations, *Extra Higgs boson searches at the LHC*, in the proceedings of the *12th International Workshop on the CKM Unitarity Triangle*, Santiago de Compostela, Spain, September 18–22 (2023) [[arXiv:2404.03571](#)] [[INSPIRE](#)].
- [51] U. Haisch, J.F. Kamenik, A. Malinauskas and M. Spira, *Collider constraints on light pseudoscalars*, *JHEP* **03** (2018) 178 [[arXiv:1802.02156](#)] [[INSPIRE](#)].
- [52] CMS collaboration, *Search for exotic Higgs boson decays  $H \rightarrow \mathcal{A}\mathcal{A} \rightarrow 4\gamma$  with events containing two merged diphotons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. Lett.* **131** (2023) 101801 [[arXiv:2209.06197](#)] [[INSPIRE](#)].
- [53] ATLAS collaboration, *Digluon Tagging using  $\sqrt{s} = 13$  TeV pp Collisions in the ATLAS Detector*, [ATL-PHYS-PUB-2021-027](#), CERN, Geneva (2021).
- [54] ATLAS collaboration, *DeXTer: Deep Sets based Neural Networks for Low- $p_T X \rightarrow b\bar{b}$  Identification in ATLAS*, [ATL-PHYS-PUB-2022-042](#), CERN, Geneva (2022).
- [55] E. Bols et al., *Jet Flavour Classification Using DeepJet*, *2020 JINST* **15** P12012 [[arXiv:2008.10519](#)] [[INSPIRE](#)].
- [56] CMS collaboration, *A new calibration method for charm jet identification validated with proton-proton collision events at  $\sqrt{s} = 13$  TeV*, *2022 JINST* **17** P03014 [[arXiv:2111.03027](#)] [[INSPIRE](#)].
- [57] K. Lassila-Perini, C. Lange, E.C. Jarrin and M. Bellis, *Using CMS Open Data in research — challenges and directions*, *EPJ Web Conf.* **251** (2021) 01004 [[arXiv:2106.05726](#)] [[INSPIRE](#)].
- [58] CERN. Open Data Portal. <https://opendata.cern.ch> [accessed: 31-Mar-2025].

- [59] CMS collaboration, *The CMS Experiment at the CERN LHC, 2008 JINST* **3** S08004 [INSPIRE].
- [60] CMS collaboration, *CMS TriDAS project, CERN-LHCC-2000-038* CERN, Geneva (2000).
- [61] CMS collaboration, *Performance of the CMS Level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV, 2020 JINST* **15** P10017 [arXiv:2006.10165] [INSPIRE].
- [62] CMS collaboration, *CMS The TriDAS Project, CERN-LHCC-2002-026*, CERN, Geneva (2002).
- [63] CMS collaboration, *The CMS trigger system, 2017 JINST* **12** P01020 [arXiv:1609.02366] [INSPIRE].
- [64] CMS collaboration, *2020 CMS data preservation, re-use and open access policy*, CERN Open Data Portal [DOI:10.7483/OPENDATA.CMS.1BNU.8V1W].
- [65] CMS collaboration, *SingleMuon primary dataset in NANO AOD format from RunG of 2016 (/SingleMuon/Run2016G-UL2016\_MiniAODv2\_NanoAODv9-v1/NANO AOD)* CERN Open Data Portal [DOI:10.7483/OPENDATA.CMS.DM27.QUP0].
- [66] CMS collaboration, *SingleMuon primary dataset in NANO AOD format from RunH of 2016 (/SingleMuon/Run2016H-UL2016\_MiniAODv2\_NanoAODv9-v1/NANO AOD)* CERN Open Data Portal [DOI:10.7483/OPENDATA.CMS.4BUS.64MW].
- [67] T. Sjöstrand et al., *An introduction to PYTHIA 8.2, Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].
- [68] CMS collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements, Eur. Phys. J. C* **76** (2016) 155 [arXiv:1512.00815] [INSPIRE].
- [69] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].
- [70] M.L. Mangano, M. Moretti, F. Piccinini and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions, JHEP* **01** (2007) 013 [hep-ph/0611129] [INSPIRE].
- [71] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, JHEP* **11** (2007) 070 [arXiv:0709.2092] [INSPIRE].
- [72] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, JHEP* **06** (2010) 043 [arXiv:1002.2581] [INSPIRE].
- [73] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ, Phys. Rev. D* **86** (2012) 094034 [arXiv:1208.5967] [INSPIRE].
- [74] M. Czakon and A. Mitov, *Top++: A program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders, Comput. Phys. Commun.* **185** (2014) 2930 [arXiv:1112.5675] [INSPIRE].
- [75] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a  $W^-$  or  $H^-$ , Phys. Rev. D* **82** (2010) 054018 [arXiv:1005.4451] [INSPIRE].
- [76] J.M. Campbell, R.K. Ellis and C. Williams, *Vector Boson Pair Production at the LHC, JHEP* **07** (2011) 018 [arXiv:1105.0020] [INSPIRE].
- [77] NNPDF collaboration, *Parton distributions from high-precision collider data, Eur. Phys. J. C* **77** (2017) 663 [arXiv:1706.00428] [INSPIRE].
- [78] GEANT4 collaboration, *GEANT4 — A Simulation Toolkit, Nucl. Instrum. Meth. A* **506** (2003) 250 [INSPIRE].

- [79] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-15To20\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.9GKI.40VV](https://doi.org/10.7483/OPENDATA.CMS.9GKI.40VV)].
- [80] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-20To30\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.8S2B.M60D](https://doi.org/10.7483/OPENDATA.CMS.8S2B.M60D)].
- [81] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-30To50\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.8HSM.3T9K](https://doi.org/10.7483/OPENDATA.CMS.8HSM.3T9K)].
- [82] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-50To80\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.708Z.2F9M](https://doi.org/10.7483/OPENDATA.CMS.708Z.2F9M)].
- [83] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-80To120\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [DOI:10.7483/OPENDATA.CMS.L7RP.3B5E](https://doi.org/10.7483/OPENDATA.CMS.L7RP.3B5E).
- [84] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-120To170\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.TT8M.UMOU](https://doi.org/10.7483/OPENDATA.CMS.TT8M.UMOU)].
- [85] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-170To300\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.98WN.CHIA](https://doi.org/10.7483/OPENDATA.CMS.98WN.CHIA)].
- [86] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-300To470\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.ZTQK.M522](https://doi.org/10.7483/OPENDATA.CMS.ZTQK.M522)].
- [87] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-470To600\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.ILVF.06MZ](https://doi.org/10.7483/OPENDATA.CMS.ILVF.06MZ)].
- [88] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-600To800\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.XZSW.EYQ6](https://doi.org/10.7483/OPENDATA.CMS.XZSW.EYQ6)].
- [89] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-800To1000\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.BZDK.2RTJ](https://doi.org/10.7483/OPENDATA.CMS.BZDK.2RTJ)].
- [90] CMS collaboration, *Simulated dataset*  
*QCD\_Pt-1000\_MuEnrichedPt5\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.PFJR.4L2J](https://doi.org/10.7483/OPENDATA.CMS.PFJR.4L2J)].
- [91] CMS collaboration, *Simulated dataset*  
*DY1jToLL\_M-1to10\_Pt-0to70\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.JVXW.GPCL](https://doi.org/10.7483/OPENDATA.CMS.JVXW.GPCL)].
- [92] CMS collaboration, *Simulated dataset*  
*DY1jToLL\_M-1to10\_Pt-70to100\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.MR2R.CDIO](https://doi.org/10.7483/OPENDATA.CMS.MR2R.CDIO)].

- [93] CMS collaboration, *Simulated dataset DY1jToLL\_M-1to10\_Pt-100to200\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.EGIQ.8DQW](https://doi.org/10.7483/OPENDATA.CMS.EGIQ.8DQW)].
- [94] CMS collaboration (2024), *Simulated dataset DY1jToLL\_M-1to10\_Pt-200to400\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.28UM.TM11](https://doi.org/10.7483/OPENDATA.CMS.28UM.TM11)].
- [95] CMS collaboration, *Simulated dataset DY1jToLL\_M-1to10\_Pt-400to600\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.OA9K.VCT7](https://doi.org/10.7483/OPENDATA.CMS.OA9K.VCT7)].
- [96] CMS collaboration, *Simulated dataset DY1jToLL\_M-1to10\_Pt-600toInf\_TuneCP5\_13TeV-madgraph-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.3JOR.RDHF](https://doi.org/10.7483/OPENDATA.CMS.3JOR.RDHF)].
- [97] CMS collaboration, *Simulated dataset DYJetsToLL\_M-10to50\_TuneCP5\_13TeV-madgraphMLM-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.AWJC.QQRN](https://doi.org/10.7483/OPENDATA.CMS.AWJC.QQRN)].
- [98] CMS collaboration, *Simulated dataset DYJetsToLL\_M-50\_TuneCP5\_13TeV-madgraphMLM-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.CRNB.POY1](https://doi.org/10.7483/OPENDATA.CMS.CRNB.POY1)].
- [99] CMS collaboration, *Simulated dataset TTTo2L2Nu\_TuneCP5\_13TeV-powheg-pythia8 in NANOADSIM in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.4RTG.JPI2](https://doi.org/10.7483/OPENDATA.CMS.4RTG.JPI2)].
- [100] CMS collaboration, *Simulated dataset ST\_tW\_Dilept\_5f\_DR\_TuneCP5\_13TeV-amcatnlo-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.HHES.TU07](https://doi.org/10.7483/OPENDATA.CMS.HHES.TU07)].
- [101] CMS] collaboration, *Simulated dataset WW\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.57IH.XPRE](https://doi.org/10.7483/OPENDATA.CMS.57IH.XPRE)].
- [102] CMS collaboration, *Simulated dataset WZ\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data* CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.1H64.40FS](https://doi.org/10.7483/OPENDATA.CMS.1H64.40FS)].
- [103] CMS collaboration, *Simulated dataset ZZ\_TuneCP5\_13TeV-pythia8 in NANOADSIM format for 2016 collision data*, CERN Open Data Portal [[DOI:10.7483/OPENDATA.CMS.WMBQ.G35Q](https://doi.org/10.7483/OPENDATA.CMS.WMBQ.G35Q)].
- [104] G. Bozzi, S. Catani, D. de Florian and M. Grazzini, *Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC*, *Nucl. Phys. B* **737** (2006) 73 [[hep-ph/0508068](https://arxiv.org/abs/hep-ph/0508068)] [[INSPIRE](https://inspirehep.net/literature/468097)].
- [105] CMS collaboration, *CMS Open Data Guide*, <https://cms-opendata-guide.web.cern.ch> [accessed: 31-Mar-2025].
- [106] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, *Nucl. Instrum. Meth. A* **389** (1997) 81 [[INSPIRE](https://inspirehep.net/literature/452506)].
- [107] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *2017 JINST* **12** P10003 [[arXiv:1706.04965](https://arxiv.org/abs/1706.04965)] [[INSPIRE](https://inspirehep.net/literature/1706049)].

- [108] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV*, 2018 *JINST* **13** P06015 [[arXiv:1804.04528](#)] [[INSPIRE](#)].
- [109] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [110] CMS collaboration, *Pileup mitigation at CMS in 13 TeV data*, 2020 *JINST* **15** P09018 [[arXiv:2003.00503](#)] [[INSPIRE](#)].
- [111] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, 2017 *JINST* **12** P02014 [[arXiv:1607.03663](#)] [[INSPIRE](#)].
- [112] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, 2018 *JINST* **13** P05011 [[arXiv:1712.07158](#)] [[INSPIRE](#)].
- [113] LHC HIGGS CROSS SECTION WORKING GROUP collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, *CERN Yellow Rep. Monogr.* **2** (2017) 1 [[arXiv:1610.07922](#)] [[INSPIRE](#)].
- [114] J. Gaiser et al., *Charmonium Spectroscopy from Inclusive psi-prime and J/psi Radiative Decays*, *Phys. Rev. D* **34** (1986) 711 [[INSPIRE](#)].
- [115] W. Verkerke and D.P. Kirkby, *The RooFit toolkit for data modeling*, *eConf C* **0303241** (2003) MOLT007 [[physics/0306116](#)] [[INSPIRE](#)].
- [116] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01 [[INSPIRE](#)].
- [117] CMS collaboration, *The CMS Statistical Analysis and Combination Tool: Combine*, *Comput. Softw. Big Sci.* **8** (2024) 19 [[arXiv:2404.06614](#)] [[INSPIRE](#)].
- [118] CMS collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, *CMS-PAS-LUM-17-001*, CERN, Geneva (2017).
- [119] CMS collaboration, *Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13$  TeV*, *JHEP* **07** (2018) 161 [[arXiv:1802.02613](#)] [[INSPIRE](#)].
- [120] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435 [[hep-ex/9902006](#)] [[INSPIRE](#)].
- [121] A.L. Read, *Presentation of search results: The  $CL_s$  technique*, *J. Phys. G* **28** (2002) 2693 [[INSPIRE](#)].