

DRAFT

CMS Paper

The content of this note is intended for CMS internal use and distribution only

2025/04/20

Archive Hash: 32611a6

Archive Date: 2023/09/28

Search for dark matter produced in association with a resonant bottom quark pair in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$

The CMS Collaboration

Abstract

A search for dark matter produced in association with a resonant bottom quark pair has been performed in proton-proton collisions at a center-of-mass energy of 13 TeV collected with the CMS detector during the 2016–2018 data-taking period at the CERN LHC. The analyzed data sample corresponds to an integrated luminosity of 138 fb^{-1} . The results are interpreted in terms of a novel theoretical model of dark matter production that, together with a spin-1 gauge boson mediator, predicts the existence of a Higgs-boson-like particle in the dark sector (i.e., a dark Higgs boson). The model simultaneously provides a mechanism to generate the masses of particles in the dark sector and a new annihilation channel that helps relax the constraints on dark matter relic abundance. If such a dark Higgs boson mixes with the standard model Higgs boson, its decay into a bottom quark pair can be identified in hadronic jets. This search focuses on final states where the dark Higgs boson is produced in association with the dark matter mediator. It gives rise to an experimental signature with a large missing transverse momentum. Limits at the 95% confidence level on the signal strength for dark Higgs boson mass hypotheses below 160 GeV are set for the first time with CMS data. Values of the mediator mass up to 2.5–4.5 TeV are excluded depending on the dark Higgs boson mass.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:	M. Cremonesi, A. Das, M. Donega, S. Eisenberger, E. Ertorer, A. Hall, M. Hildreth, B. Jayatilaka, J. Lee, N. Macilla, M. Marchegiani, C.-S. Moon, I. Pedraza, N. Smith, T. Tomei, D. Valsecchi, R. Wallny, M. Wassmer1, Z. Ye
PDFTitle:	Search for Dark Matter Produced in Association with a Resonant Bottom-Quark Pair
PDFSubject:	CMS
PDFKeywords:	CMS, physics, software, computing

Please also verify that the abstract does not use any user defined symbols

1 Introduction

The predictions of the standard model (SM) of particle physics have been confirmed by decades of experiments. Despite these successes, the SM is still not able to explain phenomena such as the existence of dark matter (DM). While astrophysical observations have established that most of the matter in the Universe is composed of DM [1], details of its nature remain elusive.

One theoretically attractive model of DM is that of a thermally produced weakly interacting massive particle (WIMP). The existence of such a particle, with the right mass and couplings, could explain the abundance of DM in the universe, as well as many of the observed phenomena commonly ascribed to DM [2]. If non-gravitational interactions exist between DM and SM particles, then the new interaction would imply the existence of a new mediator, and the Large Hadron Collider (LHC) would have the unique possibility to directly produce it along with the DM particles, and study their properties.

Simplified models of DM production at the LHC [3] have become increasingly popular in recent years. These models predict that the pair production of DM particles in hadron collisions proceeds through a spin-0 or spin-1 bosonic mediator produced in the s-channel. Such a mediator is then accompanied by some other visible SM particle, often emitted as initial-state radiation (ISR). This gives rise to experimental signatures where the mediator decays into weakly interacting DM particles, appearing as an imbalance in the transverse momentum. Such signatures are commonly referred to as “mono-X”, where X denotes either the SM particle produced in association with DM (such as a monophoton or mono-W/Z) or its detector manifestation (such as a monojet).

Among these, monojet final states – where a gluon or quark is emitted as initial-state radiation and appears in the detector as a hadronic jet – offer a favorable topology due to the high rate of radiation of quarks and gluons in the initial state. Monojet searches at the LHC [4, 5] have strongly constrained the DM parameter space, in models where DM relic particles in the cosmos would annihilate directly into final-state SM particles. The tension with astrophysical measurements of the abundance of DM is relaxed if the DM particles are not the lightest dark sector particles, leading to new annihilation channels.

The theoretical framework can therefore be extended by models in which, together with a spin-1 gauge boson Z' , a new complex Higgs field is introduced, whose vacuum expectation value spontaneously breaks the gauge symmetry in the dark sector [6], giving rise to a new physical “dark” Higgs boson H_D . If the H_D boson is sufficiently light, DM particles can annihilate into a pair of H_D bosons. This new annihilation channel would allow the model to easily match the observed relic abundance [7–9]. In this model, the DM particle χ is taken to be a Majorana fermion that couples axially to the gauge boson Z' . The Z' boson also has a vector-like coupling with SM quarks. The relevant part of the spin-1 sector of the model Lagrangian is:

$$\mathcal{L}_{\text{spin-1}} \supset -g_\chi Z'_\mu \bar{\chi} \gamma^\mu \gamma^5 \chi - g_q Z'_\mu \sum_q \bar{q} \gamma^\mu q,$$

where g_χ is the coupling between the Z' mediator and the χ particles, while g_q is the coupling between the Z' mediator and the SM quarks. These two parameters are set to 1.0 and 0.25, respectively, in accordance with the LHC Dark Matter Working Group recommendations [10].

As the lightest state in the dark sector, the H_D boson does not decay into χ particles, but it can decay into visible SM particles by mixing with the SM Higgs boson (H) [8, 11]. For this reason, the decay into a pair of b quarks is expected to be dominant for H_D bosons with masses below 135 GeV, and it is significant for masses up to 160 GeV. The mixing angle θ_h between the H_D

44 boson and the SM H boson is set to 0.01, a value that is large enough to ensure prompt decay
 45 of the H_D boson while small enough to have no observable effect on the couplings of the SM
 46 Higgs boson [12].

47 In this paper, we present a search for DM in events where χ particles are produced in associa-
 48 tion with a H_D boson decaying into a pair of b quarks. The production mechanism is shown in
 49 Fig. 1. A Z' boson is produced by a quark-antiquark interaction in the initial state. It radiates a
 50 H_D boson via dark-Higgsstrahlung before decaying into a pair of χ particles.

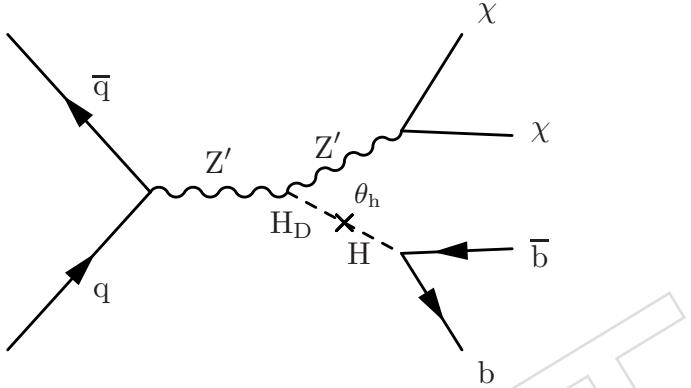


Figure 1: Feynman diagram for the associated production of a H_D boson and χ particles. The interaction with SM quarks is mediated by a Z' boson, and the H_D boson mixes with the SM Higgs boson through the θ_h mixing angle. In this paper we focus on the decay of the H_D boson into a pair of b quarks, which is dominant at lower masses.

51 Searches for H_D bosons produced in association with DM have already been performed by
 52 the ATLAS [13] and CMS [14] collaborations. These searches focus on heavier H_D boson mass
 53 hypotheses, larger than 160 GeV. For a H_D boson of such a mass, the decay into a pair of W
 54 bosons is dominant. The ATLAS Collaboration has also recently published a search for lower-
 55 mass H_D bosons decaying into a pair of b quarks [15]. In this paper, we describe a similar
 56 search which uses the full dataset collected by the CMS experiment at a center-of-mass energy
 57 of 13 TeV during the 2016–2018 data taking period, corresponding to an integrated luminosity
 58 of 138 fb^{-1} .

59 2 The CMS detector and event reconstruction

60 The CMS apparatus [16, 17] is a multipurpose, nearly hermetic detector, designed to trigger
 61 on [18–20] and identify electrons, muons, photons, and (charged and neutral) hadrons [21–23].
 62 Its central feature is a superconducting solenoid of 6 m internal diameter, providing a mag-
 63 netic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead
 64 tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron cal-
 65 orimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters
 66 extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are
 67 reconstructed using gas-ionization detectors embedded in the steel flux-return yoke outside
 68 the solenoid. More detailed descriptions of the CMS detector, together with a definition of the
 69 coordinate system used and the relevant kinematic variables, can be found in Refs. [16, 17].

70 The silicon tracker used in 2016 measured charged particles in the range $|\eta| < 2.5$. For non-
 71 isolated particles of $1 < p_T < 10 \text{ GeV}$ and $|\eta| < 1.4$, the track resolutions were typically 1.5% in
 72 p_T and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [23]. At the begin-
 73 ning of 2017, a new pixel detector was installed [24]; the upgraded tracker measured particles

⁷⁴ up to $|\eta| = 3.0$ with typical resolutions of 1.5% in p_T and 20–75 μm in the transverse impact
⁷⁵ parameter [25] for non-isolated particles of $1 < p_T < 10 \text{ GeV}$. According to simulation studies
⁷⁶ [26], similar improvements are expected in the longitudinal direction. The primary vertex
⁷⁷ (PV) is taken to be the vertex corresponding to the hardest scattering in the event, evaluated
⁷⁸ using tracking information alone, as described in Section 9.4.1 of Ref. [27].

⁷⁹ In the region $|\eta| < 1.74$, the HCAL cells have widths of 0.087 in pseudorapidity and 0.087 in azim-
⁸⁰ uth (ϕ). In the η - ϕ plane, and for $|\eta| < 1.48$, HCAL cells map to 5×5 arrays of ECAL crystals
⁸¹ to form calorimeter towers projecting radially outwards from close to the nominal interaction
⁸² point. For $|\eta| > 1.74$, the coverage of the towers increases progressively to a maximum of 0.174
⁸³ in $\Delta\eta$ and $\Delta\phi$. The forward hadron (HF) calorimeter uses steel as an absorber and quartz fibers
⁸⁴ as the sensitive material. The two halves of the HF are located 11.2 m from the interaction re-
⁸⁵ gion, one at each end, and together they provide coverage in the range $3.0 < |\eta| < 5.2$. They
⁸⁶ also serve as luminosity monitors.

⁸⁷ Events of interest are selected using a two-tier trigger system. The first level (L1), composed
⁸⁸ of custom hardware processors, uses information from the calorimeters and muon detectors to
⁸⁹ select events at a rate of around 100 kHz within a fixed latency of $4 \mu\text{s}$ [18]. The second level,
⁹⁰ known as the high-level trigger (HLT), consists of a farm of processors running a version of the
⁹¹ full event reconstruction software optimized for fast processing, and reduces the event rate to
⁹² a few kHz before data storage [19, 20].

⁹³ A particle-flow (PF) algorithm [28] aims to reconstruct and identify each individual particle in
⁹⁴ an event, with an optimized combination of information from the various elements of the CMS
⁹⁵ detector. In this process, the identification of the PF candidate type (photon, electron, muon,
⁹⁶ and charged and neutral hadrons) plays an important role in the determination of the particle
⁹⁷ direction and energy. The energy of the photons is obtained from the ECAL measurement. The
⁹⁸ energy of electrons is determined from a combination of the electron momentum at the primary
⁹⁹ interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster,
¹⁰⁰ and the energy sum of all bremsstrahlung photons spatially compatible with originating from
¹⁰¹ the electron track. The energy of muons is obtained from the curvature of the corresponding
¹⁰² track. The energy of charged hadrons is determined from a combination of their momentum
¹⁰³ measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for
¹⁰⁴ the response function of the calorimeters to hadronic showers. Finally, the energy of neutral
¹⁰⁵ hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

¹⁰⁶ In this search, electrons (photons) are required to have $p_T > 10$ (15) GeV and $|\eta| < 2.5$. Muons
¹⁰⁷ are required to have $p_T > 10 \text{ GeV}$ and $|\eta| < 2.4$. All leptons and photons are required to be
¹⁰⁸ isolated. Isolation is calculated by imposing thresholds on the energy of PF candidates within a
¹⁰⁹ certain distance $\Delta R = \sqrt{(\Delta\phi^2 + \Delta\eta^2)}$ with respect to the lepton/photon. Additional selection
¹¹⁰ criteria are applied to define “loose” (“veto”) electrons (muons and photons) [21, 22], which
¹¹¹ are used to reject unwanted events. Similarly, “tight” leptons/photons are defined and used to
¹¹² select events in control data samples.

¹¹³ Hadronically decaying tau leptons are required to pass identification criteria using the hadron-
¹¹⁴ plus-strips algorithm [29]. In addition, a new algorithm for the identification of hadronic tau
¹¹⁵ lepton decays, called DEEPTAU [30], is used. The DEEPTAU algorithm is based on multi-
¹¹⁶ classification and provides classifiers to discriminate genuine hadronic tau lepton decays from
¹¹⁷ jets, electrons, and muons. Nonetheless, the overlap with electrons and muons is accounted for
¹¹⁸ by removing tau leptons that are within a distance $\Delta R < 0.4$ around a well reconstructed and
¹¹⁹ isolated electron or muon.

For each event, hadronic jets are clustered from the PF candidates using the infrared and collinear safe anti- k_T algorithm [31, 32] with a distance parameter of 0.4 or 1.5. Depending on the respective distance parameter, these jets are referred to as “AK4” or “AK15” jets. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_T spectrum and detector acceptance. Jet energy corrections are derived from simulation to bring the measured response of jets to that of particle-level jets on average. In situ measurements of momentum balance in dijet, photon + jet, Z + jet, and multijet events are used to account for any residual differences in the jet energy scale between data and simulation [33]. The jet energy resolution typically amounts to 15—20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [33]. Additional selection criteria are applied to each jet to remove jets potentially dominated by anomalous contributions from various subdetector components or reconstruction failures.

Additional proton-proton interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect in the AK4 jets, charged particles identified to be originating from pileup vertices are discarded, and an offset correction is applied to correct for remaining contributions. For AK15 jets, the pileup per particle identification algorithm (PUPPI) [34, 35] is used to mitigate the effect of pileup at the reconstructed particle level, making use of local shape information, event pileup properties, and tracking information. A local shape variable is defined, which distinguishes between collinear and soft diffuse distributions of other particles surrounding the particle under consideration. The former is attributed to particles originating from the hard scatter, and the latter is attributed to particles originating from pileup interactions. Charged particles identified to be originating from pileup vertices are discarded. For each neutral particle, a local shape variable is computed using the surrounding charged particles compatible with the primary vertex within the tracker acceptance ($|\eta| < 2.5$), and using both charged and neutral particles in the region outside of the tracker coverage. The momenta of the neutral particles are then rescaled according to their probability to originate from the primary interaction vertex deduced from the local shape variable, superseding the need for jet-based pileup corrections [34].

The AK4 jets used in this search are further required to have a p_T larger than 30 GeV and $|\eta| < 2.5$. Jets with $\Delta R < 0.4$ with respect to a well identified and isolated lepton or photon are removed. To identify jets originated by the hadronization of b quarks (hereafter referred to as “b jets”), the DEEPJET algorithm [36] is employed. A loose working point is used, defined for each year of data-taking as the minimum requirement in the DEEPJET discriminator distribution, which returns a 10% rate of misidentifying a jet originated by a light-flavor quark. The loose working point corresponds to an efficiency of correctly identifying jets originated by b quarks (i.e., b-tagging efficiency) of 90–95%, depending on the p_T of the AK4 jet.

The AK15 jets used in this search are further required to have a p_T larger than 160 GeV and $|\eta| < 2.4$. Jets with $\Delta R < 1.5$ with respect to a well-identified and isolated lepton or photon are removed. The modified mass drop tagger algorithm [37, 38], also known as the soft-drop (SD) algorithm, with the angular exponent $\beta = 0$, soft cutoff threshold $z_{\text{cut}} < 0.1$, and characteristic radius $R_0 = 1.5$ [39], is applied to remove soft, wide-angle radiation from the jet. To identify AK15 jets that are consistent with the hadronization of a $b\bar{b}$ pair from the decay of a boosted massive resonance, the DEEPAK15 algorithm [40] is used. More specifically, the “mass-decorrelated” (MD) version of DEEPAK15 is employed. In this variant, an adversarial training is performed in which a second neural network is made to extract the AK15 jet mass from the output of the DEEPAK15 graph neural network. A good performance of this second network yields to a penalty on the joint cost function of the two networks. Therefore, this

method optimizes the ability to correctly identify the origin of an AK15 jet while systematically decorrelating the output score from the AK15 jet mass. This approach avoids shaping the AK15 jet mass distribution in background events. Since the strategy for this search relies on the AK15 jet mass shape for background estimation, the MD version of the tagger offers the best option. Identification working points are defined for each year of data-taking and optimized for this specific search.

The missing transverse momentum vector \vec{p}_T^{miss} is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as p_T^{miss} . The \vec{p}_T^{miss} is modified to account for corrections to the energy scale and resolution of the reconstructed AK4 jets in the event [41]. Anomalous high- p_T^{miss} events can be due to a variety of reconstruction failures, detector malfunctions, or non-collision backgrounds. Such events are rejected by dedicated filters that are designed to eliminate more than 85–90% of spurious high- p_T^{miss} events with a signal efficiency exceeding 99.9% [41].

Together with the AK15 jet mass, the hadronic recoil, called U, is also used to distinguish the signal from the backgrounds. It represents the total transverse momentum of all the non-hadronic particles in each event. The signal events in this search contain only jets and no other reconstructed candidates, therefore U is equivalent to p_T^{miss} of the event. For the leading background processes, identified in Sect. 4, U corresponds to the p_T of a vector boson. In those data samples of this search where no lepton is required, the lepton from the vector boson decay is missed and U is once again equivalent to the p_T^{miss} of the event. In data samples where a lepton is required in the final state, the lepton from the vector-boson decay is identified and U is equivalent to the magnitude of the vector sum of \vec{p}_T^{miss} and the lepton \vec{p}_T .

3 Simulated samples

Samples of Monte Carlo (MC) simulated events are used to predict the signal and background contributions. In all cases, parton showering, hadronization, and underlying event properties are modeled using PYTHIA [42] version 8.202 or later with the underlying event tune CUETP8M1 or CP5 [43], based on the year of data taking. The simulation of the interactions between the particles and the CMS detector is based on GEANT4 [44]. The NNPDF 3.0 next-to-next-to-leading order (NNLO) [45] and NNPDF 3.1 NNLO [46] parton distribution functions (PDFs) are used for the generation of all samples based on the year of data collection. The same reconstruction algorithms used for the data are applied to the simulated samples.

For the associated production of SM vector bosons and jets ($V + \text{jets}$ production), predictions with up to two partons in the final state are obtained at leading order (LO) in QCD using MADGRAPH5_aMC@NLO [47] with the MLM matching scheme [48] between the jets from the calculations of the matrix elements and the parton shower. Samples of events with top quark pairs ($t\bar{t}$ production) are generated at next-to-leading (NLO) in QCD with up to two additional partons in the matrix element calculations using MADGRAPH5_aMC@NLO and the FxFx jet matching scheme [49]. Their cross sections are normalized to the inclusive cross section of $t\bar{t}$ production at NNLO in QCD [50]. Events with electroweakly produced single top quarks (single top production) are simulated using POWHEG 2.0 [51, 52] and normalized to the inclusive cross section calculated at NLO in QCD [53, 54]. The associated production of vector bosons (VV production) is simulated at NLO in QCD using PYTHIA, and normalized to the cross sections at NNLO precision for WW production [55], and at NLO precision for WZ and ZZ production [56]. Several production mechanisms of SM H bosons decaying into a pair of b quarks ($H \rightarrow b\bar{b}$ production) are also produced at LO with the POWHEG generator. Samples of QCD multijet production events are generated at LO using MADGRAPH5_aMC@NLO.

214 Simulated samples of H_D boson production are generated with MADGRAPH5_aMC@NLO [47]
 215 at LO, including up to one additional parton in the matrix element calculations, with the MLM
 216 matching scheme. Separate samples are generated for different mass hypotheses for the Z'
 217 mediator, χ particles, and H_D bosons.

218 4 Event selection

219 The signal targeted in this analysis displays a large U and an AK15 jet that is identified as
 220 originating from a $H_D \rightarrow b\bar{b}$ decay. Events with these characteristics comprise the signal
 221 region (SR).

222 Events in the SR are collected by trigger selections based on two variables. The first is p_T^{miss} ,
 223 which is calculated using all PF candidates reconstructed at the HLT except for muons. The
 224 second variable, called H_T^{miss} , is defined as the magnitude of the negative vector sum of \vec{p}_T of
 225 the hadronic jets in the event, using the AK4 jets clustered by the HLT reconstruction. The
 226 trigger selects events with both p_T^{miss} and $H_T^{\text{miss}} > 120$ GeV.

227 Events in SR are selected with a requirement of $U > 250$ GeV. The soft-drop corrected mass
 228 (m_{SD}) of the leading AK15 jet in p_T must satisfy the requirement $m_{SD} \in [40, 300]$. In order to
 229 preferentially select events where the leading AK15 jet originates from the hadronic decay of a
 230 H_D boson, the jet is further required to satisfy a minimum requirement in the DEEPAK15 score.
 231 A different threshold is used for each year of data-taking, optimized against signal sensitivity.

232 After this basic pre-selection, the main background processes in this search are from $Z(\nu\nu) +$
 233 jets (hereafter referred to as $Z + \text{jets}$), $W(e\nu)/W(\mu\nu) + \text{jets}$ (collectively referred to as $W + \text{jets}$),
 234 $t\bar{t}$ and QCD multijet production. The $Z + \text{jets}$ process is the largest background and is irre-
 235ducible. In contrast, the background from $W + \text{jets}$ and $t\bar{t}$ processes is suppressed by rejecting
 236 events if they contain a well-reconstructed and isolated lepton. For electrons, the "veto" work-
 237 ing point (defined in Sect. 2) is used in their identification, while the "loose" working point is
 238 used in the identification of muons. Events that contain a "loose" photon are also rejected. This
 239 helps to suppress electroweak backgrounds in which a photon is radiated from the initial state.
 240 The $t\bar{t}$ process is further suppressed by vetoing events with b-tagged AK4 jets that do not over-
 241 lap with the leading AK15 jet. Finally, to reject QCD multijet events with large U arising from
 242 mismeasurements of the jet momenta, a minimum requirement is imposed on the azimuthal
 243 separation between the \vec{U} direction and each AK4 jet in the event to be larger than 0.5 radians.
 244 Similarly, the azimuthal angle between the \vec{U} direction and each AK15 jet in the event must be
 245 larger than 1.5 radians.

246 To avoid events with anomalous U due to reconstruction failures of the PF algorithm, events
 247 are required to have $|p_T^{\text{miss}}(\text{PF}) - p_T^{\text{miss}}(\text{calorimeter})|/U < 0.5$, where $p_T^{\text{miss}}(\text{PF})$ refers to the
 248 standard p_T^{miss} computed from PF candidates as defined in Sec. 2, while $p_T^{\text{miss}}(\text{calorimeter})$ is
 249 the p_T^{miss} calculated using only information from the calorimeters. To mitigate mismeasurement
 250 from a non-functioning HCAL section in 2018 data, events with jets in that region or with
 251 $\phi(\vec{p}_T^{\text{miss}}) \in [-1.62, -0.62]$ at low p_T^{miss} are rejected.

252 As shown in Table 1, even after the full SR selection is applied, the selected data sample still
 253 has a large contamination from $Z + \text{jets}$, $W + \text{jets}$, and $t\bar{t}$ production. In order to predict and
 254 constrain these background processes, dedicated control regions (CRs) are used.

Table 1: Expected yields from background processes in SR. The values shown are from MC simulation and the uncertainties are statistical-only.

	2016	2017	2018
$H \rightarrow b\bar{b}$	57.6 ± 0.3	72.0 ± 0.3	83.8 ± 0.3
$Z(\rightarrow \ell\ell) + \text{jets}$	56.8 ± 2.2	43.3 ± 2.0	37.1 ± 3.0
QCD multijet	93.3 ± 25.8	154.9 ± 41.7	163.2 ± 64.6
VV	718.0 ± 17.5	623.4 ± 17.8	606.4 ± 20.8
Single top	646.0 ± 10.9	567.4 ± 12.5	614.6 ± 12.8
$t\bar{t}$	5486.5 ± 199.7	5810.7 ± 60.0	6784.2 ± 133.7
$W + \text{jets}$	3997.8 ± 38.5	2991.0 ± 40.2	2826.6 ± 50.5
$Z + \text{jets}$	7514.8 ± 29.2	7035.2 ± 33.3	6978.5 ± 38.8
Total background	18570.7 ± 208.1	17297.9 ± 92.4	18094.2 ± 163.4

255 A CR enriched in $Z + \text{jets}$ production events is identified using the same requirements that
 256 define the SR, but inverting the criterion on the DEEPAK15 score.

257 Similarly, CRs enriched in $W + \text{jets}$ production events are identified using the full SR selection
 258 criteria with the exception of the muon or electron veto. More specifically, single-muon CRs
 259 are composed of events with exactly one tight muon, and single-electron CRs are composed of
 260 events with exactly one tight electron.

261 At the HLT, events with high- p_T muons are also assigned large p_T^{miss} , therefore the same triggers
 262 based on p_T^{miss} and H_T^{miss} are also used to collect data populating the single-muon CRs. The
 263 control samples with electrons are selected using two single-electron triggers: one that requires
 264 $p_T > 27$ (2016), 35 (2017), 32 (2018) GeV, while the other requires $p_T > 105$ (2016), 115 (2017–
 265 2018) GeV. Additionally, a single-photon trigger with $p_T > 200$ GeV is used in 2017 and 2018.
 266 Single-electron triggers differ in their isolation requirements: While the lower threshold trigger
 267 requires electrons to be well isolated, the higher threshold trigger does not, which improves
 268 the efficiency at high p_T . Similarly, the single-photon trigger avoids reliance on the HLT track
 269 reconstruction and increases the overall efficiency for electrons with $p_T > 200$ GeV. During the
 270 2016 and 2017 data taking, a gradual shift in the timing of the inputs of the ECAL L1 trigger in
 271 the region $|\eta| > 2.0$ caused a specific trigger inefficiency, known as L1 pre-firing. Correction
 272 factors are computed from the data and applied to the acceptance evaluated by simulation for
 273 the 2016 and 2017 samples.

274 Although the same requirement on U that defines the SR is used in all single-lepton CRs, to
 275 suppress the larger contribution from QCD multijet events in single-electron CRs due to jets
 276 faking electrons, the p_T^{miss} in each event is required to be larger than 100 GeV. An additional
 277 set of single-lepton CRs is used in this search, populated with events that meet the aforemen-
 278 tioned single-electron or single-muon requirements, but fail the criterion on the leading AK15
 279 jet DEEPAK15 score.

280 Dedicated single-lepton CRs are also used to constrain the background from $t\bar{t}$ production.
 281 The same single-electron or single-muon selections used to define the $W + \text{jets}$ production CRs
 282 are applied, but the veto on b-tagged AK4 jets that do not overlap with the leading AK15 jet
 283 is inverted. In this case, only CRs populated with events that satisfy the leading AK15 jet
 284 DEEPAK15 score requirement are used.

285 5 Background estimation

286 Background estimation and signal extraction are performed simultaneously, using a joint maximum
 287 likelihood (ML) fit across all SR and CRs for each year. A likelihood function is con-
 288 structed to model the expected background contributions in each bin of the two-dimensional
 289 U-vs- m_{SD} variable of the SR and CRs, as well as the expected signal yield in each bin of the SR.
 290 The best fit background model, as well as the best fit signal strength modifier μ (which—for
 291 a given signal hypothesis—controls the signal normalization relative to the theoretical cross
 292 section), are obtained by maximizing a joint likelihood function of all SRs and CRs.

293 Separate approaches are adopted to estimate the dominant backgrounds from $Z + \text{jets}$, $W +$
 294 jets, and $t\bar{t}$ processes, and the subdominant backgrounds from the single top, VV, $H \rightarrow b\bar{b}$, and
 295 QCD multijet ones. The predictions for the dominant backgrounds $Z + \text{jets}$, $W + \text{jets}$, and
 296 $t\bar{t}$ in SR are based on the yield of the same processes in each bin of the CRs. The per-bin
 297 yields for these processes in SR are defined as free parameters of the likelihood function. A
 298 different set of free parameters are used for each year of data-taking. The yields in the CRs are
 299 then defined relative to these parameters by introducing a set of per-bin transfer factors. This
 300 choice of transfer factors takes into account the correlations between the $Z + \text{jets}$ and $W + \text{jets}$
 301 background contributions in all regions. In all cases, the central values of the transfer factors
 302 are obtained from the ratios of the simulated U vs. m_{SD} spectra of the respective processes in
 303 the SR to those in the CRs. The predictions for the subdominant single top, VV, $H \rightarrow b\bar{b}$, and
 304 QCD multijet backgrounds in all SR and the CRs are taken directly from simulation.

305 The likelihood method relies on the accurate predictions of the ratios between the dominant
 306 backgrounds in the SRs and CRs, as well as on the absolute normalization and shape of the
 307 U-vs.- m_{SD} distributions for the subdominant backgrounds. To achieve the most accurate pos-
 308 sible predictions for these quantities, weights are applied to each simulated event to take into
 309 account both experimental and theoretical effects that are not present in the MC samples. The
 310 experimental corrections are related to the trigger efficiencies, the identification and reconstruc-
 311 tion efficiencies of charged leptons, the efficiencies of the DEEPJET and DEEPAK15 algorithms,
 312 and the pileup distribution in simulation. Theoretical corrections are applied to the $Z + \text{jets}$
 313 and $W + \text{jets}$ processes in order to model the effects of NLO terms in the perturbative EW cor-
 314 rections [57]. Corrections are parameterized as a function of the generator-level boson p_{T} and
 315 are evaluated separately for the $W + \text{jets}$ and $Z + \text{jets}$ processes.

316 5.1 Systematic uncertainties

317 Systematic uncertainties are incorporated into the likelihood function as nuisance parameters.
 318 In the case of the $Z + \text{jets}$, $W + \text{jets}$, and $t\bar{t}$ processes, the nuisance parameters affect the values
 319 of the transfer factors in each bin of the U-vs- m_{SD} variable and thus control the ratios of the
 320 contributions from different processes, as well as the ratios of the yields in the SRs to those in
 321 various CRs. For the subdominant background processes, the yields in each bin are directly
 322 parameterized in terms of the nuisance parameters.

323 Uncertainties in the measurement of the integrated luminosity in each year of data taking are
 324 0.6–2.0% [58–60]. The uncertainties in the corrections for the L1 pre-firing effect in 2016 and
 325 2017, as well as the uncertainties in the pileup correction are of the order of 1%. The uncertain-
 326 ties in the efficiencies of reconstructing and identifying electron candidates are 1% and 2–3%,
 327 respectively. For muons, the uncertainties in the identification efficiency are 1%, with an ad-
 328 ditional 1% uncertainty in the efficiency of the isolation criteria. A systematic uncertainty for
 329 each lepton/photon veto selection has been obtained by propagating the overall uncertainties
 330 in the identification of muons, electrons, photons, and taus into the vetoed regions. While the

331 uncertainties are found to be negligible for the photon, muon and electron vetoes, a 3% un-
332 certainty in the tau veto is included. The uncertainties in the trigger efficiency are 1% for the
333 single electron trigger and 1–2% for the p_T^{miss} trigger. The uncertainty in the modeling of p_T^{miss}
334 in simulation [61] is dominated by the uncertainty in the jet energy corrections. The resulting
335 bin migration affects the acceptance of the minimum requirement in U. The resulting change in
336 rate is estimated to be 5%, and is included as a systematic uncertainty. The effect of the uncer-
337 tainties in the AK15 jet energy corrections on the jet p_T is considered. In addition, in this case,
338 the resulting bin migration affects the acceptance of the minimum requirements in the AK15
339 jet p_T . This introduces an effect on the rate of the order of 4%. The uncertainty in the DEEPJET
340 efficiency leads to a shape uncertainty applied to all processes in all regions. The uncertainty
341 in the DEEPAK15 efficiency results in a shape uncertainty applied to the signal processes in
342 SR. Uncertainties of 100% are assigned to normalization of the QCD multijet background con-
343 tributions in all regions. These uncertainties are correlated among the regions with the same
344 source of fakes: an uncertainty is applied to QCD multijet events in the SR and in the CR en-
345 riched in Z + jets production events, a separate uncertainty is applied to QCD multijet events in
346 single-muon CRs, and another uncertainty is applied to single-electron CRs. Additionally, the
347 uncertainties of 20% are assigned to the cross section of VV, H → b̄b, and DY+jets productions.
348 Similarly, 10% uncertainties in the single top and t̄t production cross sections are also assigned.
349 The theoretical uncertainties in the transfer factors related to higher-order effects in the QCD
350 and EW perturbative expansions are calculated according to the prescription given in Ref. [57],
351 and implemented as described in Ref. [62]. Bin-by-bin statistical uncertainties are incorporated
352 following the Barlow-Beeston-lite approach [63].

353 The likelihood functions obtained for the three data-taking years are combined to maximize
354 the statistical power of the search. The combination is performed by defining a combined
355 likelihood that describes all the regions in all data sets. For this purpose, the effects of all
356 theoretical uncertainties are assumed to be correlated. Most experimental uncertainties are
357 dominated by the inherent precision of auxiliary measurements specific to each dataset and
358 are thus assumed to be uncorrelated among the different data taking years. The experimental
359 uncertainties related to the determination of the integrated luminosity and to the DEEPJET
360 efficiency are partially correlated among the data taking years, which is taken into account by
361 splitting the total uncertainty into its correlated and uncorrelated components. A summary of
362 all the uncertainties considered for this analysis is reported in Table 2.

Table 2: Summary of statistical and systematic uncertainties included in the analysis. The value given for each uncertainty is the maximum value.

Source	Uncertainty
Luminosity	0.6–2%
Pileup	$\mathcal{O}(1\%)$
L1 pre-firing	$\mathcal{O}(1\%)$
p_T^{miss} trigger efficiency	1–2%
Single electron trigger efficiency	1%
Muon isolation efficiency	1%
Muon identification efficiency	1%
Electron reconstruction efficiency	1%
Electron identification efficiency	2–3%
p_T^{miss}	5%
Jet energy corrections	4%
DEEPJET efficiency	shape
DEEPAK15 efficiency	shape
VV cross section	20%
$H \rightarrow b\bar{b}$ cross section	20%
DY+jets cross section	20%
single top cross section	10%
$t\bar{t}$ cross section	10%
QCD- p_T^{miss} normalization	100%
QCD-electron normalization	100%
QCD-muon normalization	100%
Higher-order corrections	shape
Bin-by-bin statistics	shape

363 6 Results and interpretation

364 The ML fit is performed by combining the SR and CRs as well as the datasets corresponding
 365 to the years of data taking. The U-vs.- m_{SD} distributions in SR before and after the fit (“prefit”
 366 and “postfit”) for all three years combined are shown in Fig. 2. Good agreement is observed
 367 between the background-only postfit result and the data.

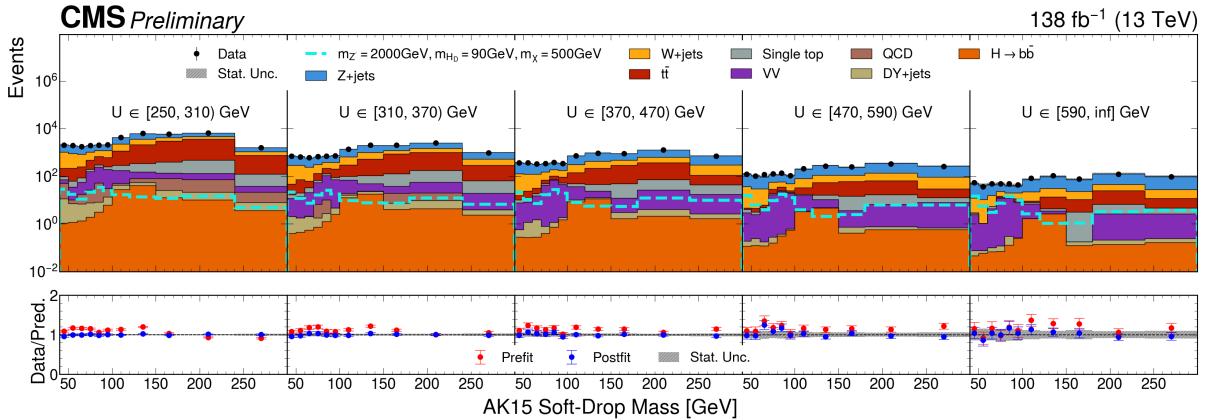


Figure 2: Postfit m_{SD} distributions in bins of U for all three years combined. Distributions in SR are shown. The top plots present stacked postfit predictions for the backgrounds superimposed on the data. The blue curve represents the postfit total background prediction, while the red curve represents the prefit one. The bottom plots present the ratio between the data (labeled as “Observed”, or “Obs”) and the background predictions (labeled as “Expected”, or “Exp”). The ratio between the data and the postfit prediction is represented by the blue dots, while the ratio between the data and the prefit prediction is represented by the red ones. Only statistical uncertainties are shown.

368 No signal is observed in the data; therefore, exclusion limits on μ are presented for different
 369 signal hypotheses. All data sets and categories are included. The exclusion limits are calculated
 370 using the CL_s criterion [64–66], and an asymptotic approximation to the distribution of the
 371 profiled likelihood ratio test statistic.

372 Exclusion limits are calculated in the two-dimensional parameter space of the DM and medi-
 373 ator masses, m_χ and $m_{Z'}$, constrained by the fact that only scenarios in which the DM particle
 374 is more massive than the H_D boson are considered. The coupling between the mediator and
 375 the SM quarks is set to a constant value of $g_q = 0.25$, and the mediator-DM coupling is set
 376 to $g_\chi = 1.0$. The resulting exclusion limits at the 95% confidence level (CL) on μ are shown
 377 in Figs. 3–8 for different hypotheses of the H_D boson mass. In the plots, darker shades corre-
 378 spond to smaller upper limits, i.e. more stringent constraints. The solid black line represents
 379 the observed 95% CL exclusion contour, while the dashed and dotted lines indicate the me-
 380 dian expected exclusion and its 68% and 95% confidence intervals, respectively. The parameter
 381 space inside the solid black boundary is excluded at the 95% CL under the model assumptions.
 382 Values of the mediator mass of up to 2.5–4.5 TeV are excluded, depending on the mass of the
 383 H_D boson.

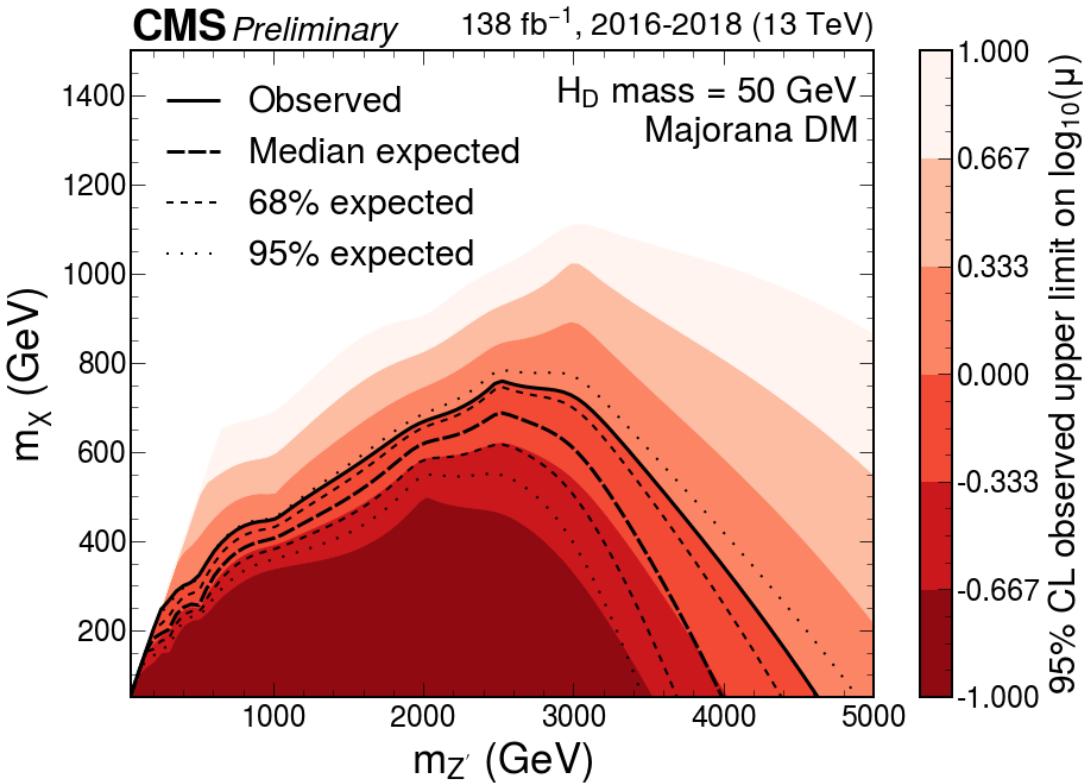


Figure 3: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 50 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

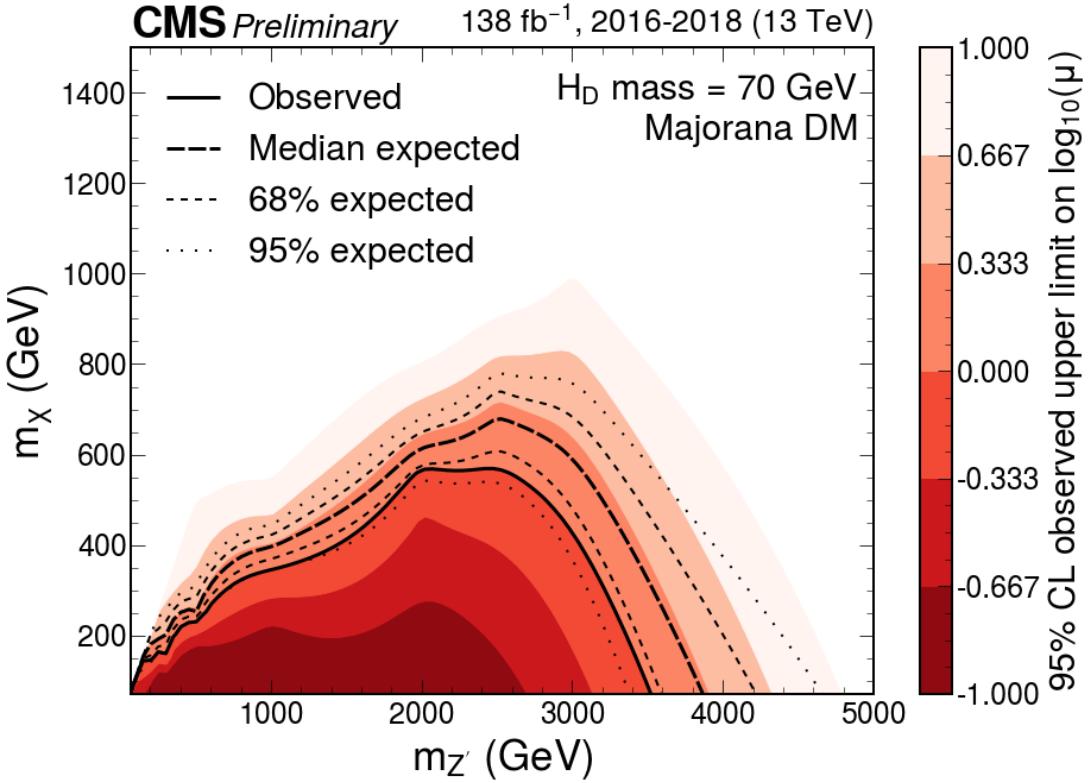


Figure 4: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 70 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

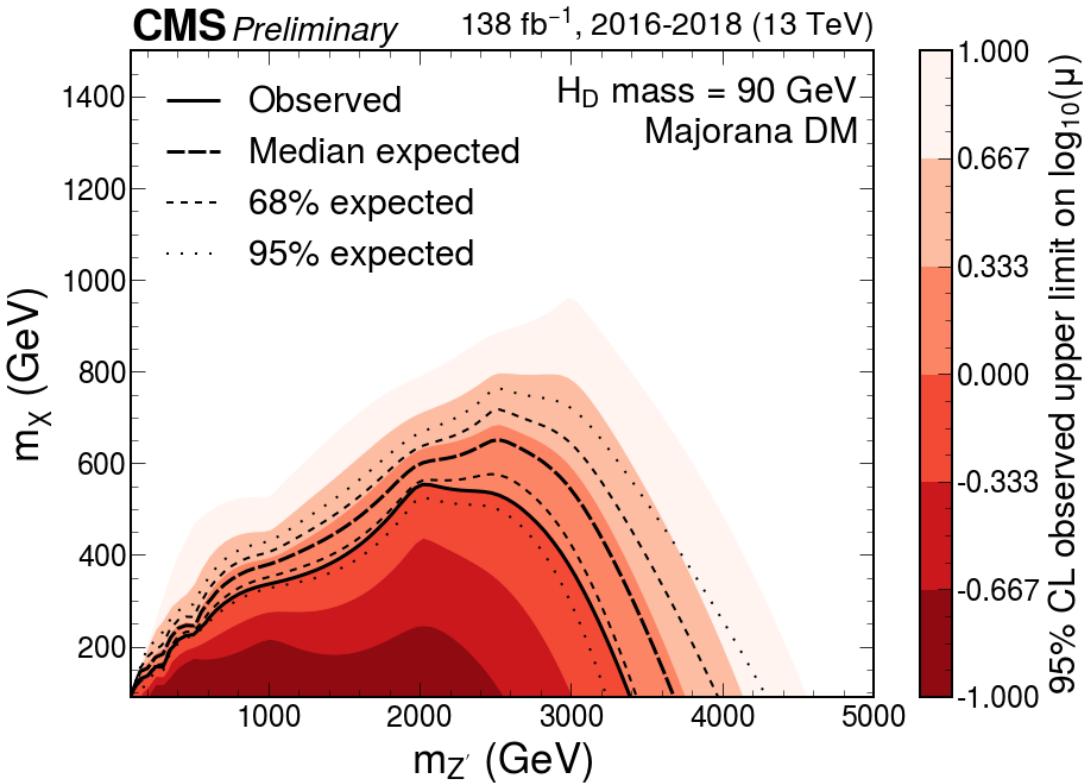


Figure 5: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 90 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

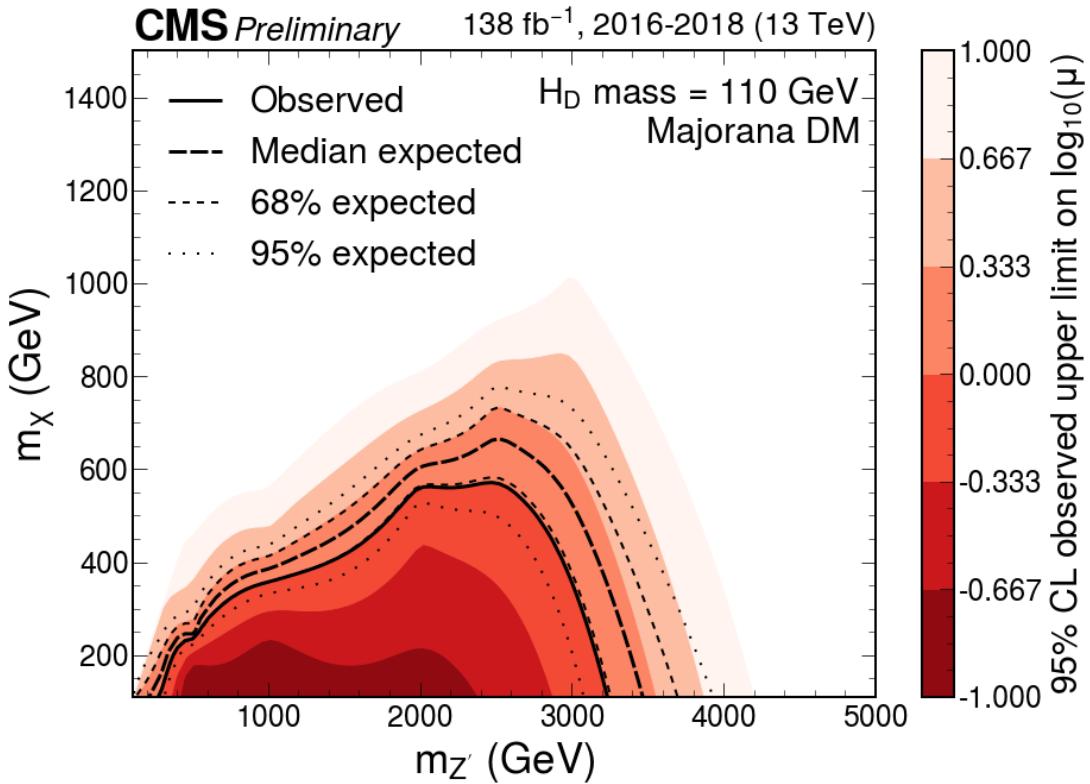


Figure 6: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 110 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

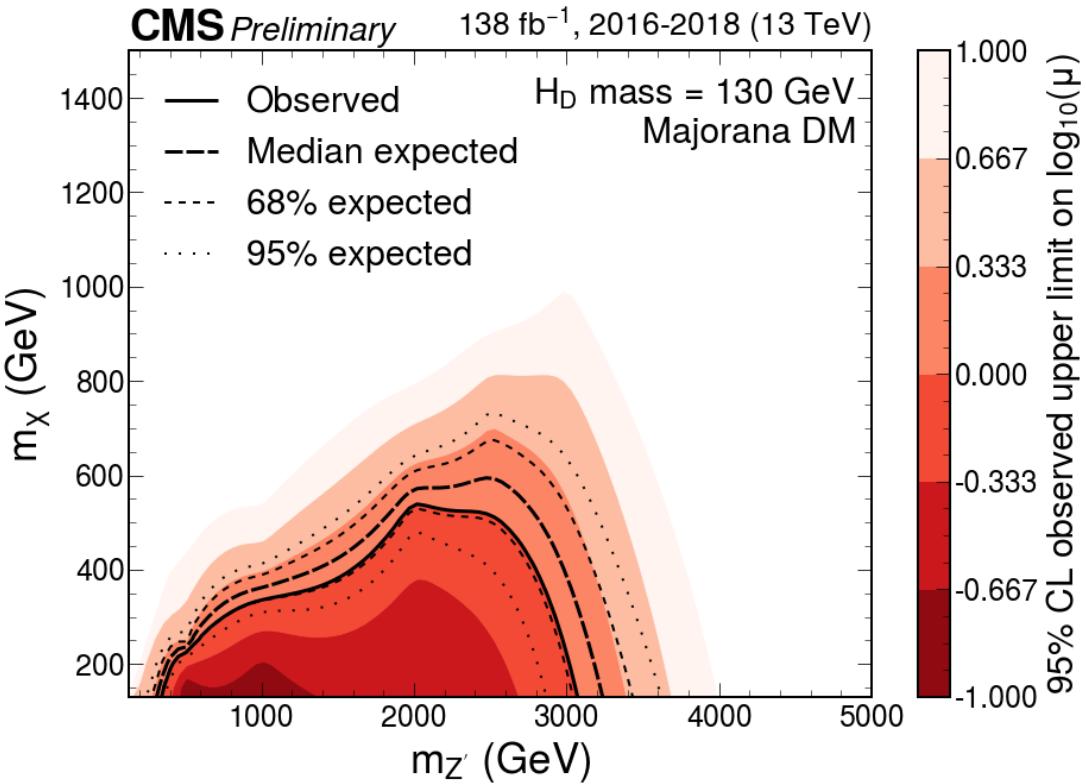


Figure 7: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 130 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

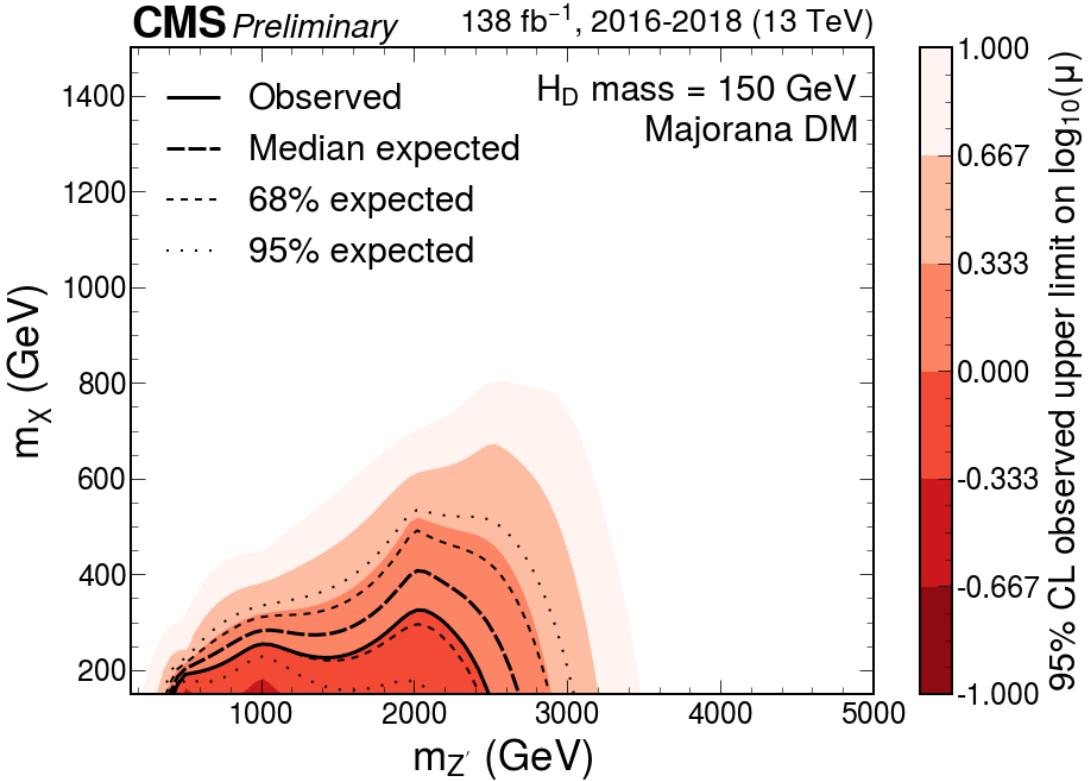


Figure 8: Expected and observed exclusion limits at 95% CL on the signal strength $\mu = \sigma/\sigma_{\text{theo}}$ as a function of $m_{Z'}$ for a H_D boson mass of 150 GeV. Only scenarios where the DM particle is more massive than H_D are considered. The black solid line indicates the exclusion boundary $\mu = 1$. Parameter combinations with larger values of μ are excluded.

384 7 Summary

385 A search for physics beyond the standard model in events with a resonant pair of b quarks
 386 and large missing transverse momentum has been presented. A data set of proton-proton
 387 collisions at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of
 388 138 fb^{-1} is analyzed. A joint maximum likelihood fit over a combination of signal and control
 389 regions is used to constrain the standard model background processes and to extract a possible
 390 signal. The result is interpreted in terms of exclusion limits at the 95% confidence level on the
 391 parameters of a model of production of a H_D boson in association with dark matter particles.
 392 Values of the mediator mass of up to 2.5–4.5 TeV are excluded, depending on the mass of the
 393 H_D boson and assuming couplings of $g_q = 0.25$ between the mediator and quarks, and $g_\chi = 1.0$
 394 between the mediator and the DM particles.

395 **Acknowledgments**

396 **References**

- 397 [1] N. Arkani-Hamed et al., “A theory of dark matter”, *Phys. Rev. D* **79** (2009) 015014,
398 doi:[10.1103/PhysRevD.79.015014](https://doi.org/10.1103/PhysRevD.79.015014).
- 399 [2] G. Bertone, D. Hooper, and J. Silk, “Particle dark matter: Evidence, candidates and
400 constraints”, *Phys. Rept.* **405** (2005) 279–390,
401 doi:[10.1016/j.physrep.2004.08.031](https://doi.org/10.1016/j.physrep.2004.08.031), arXiv:[hep-ph/0404175](https://arxiv.org/abs/hep-ph/0404175).
- 402 [3] D. Abercrombie et al., “Dark matter benchmark models for early LHC Run-2 searches:
403 report of the ATLAS/CMS dark matter forum”, 2015. arXiv:[1507.00966](https://arxiv.org/abs/1507.00966).
- 404 [4] CMS Collaboration, “Search for new particles in events with energetic jets and large
405 missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11**
406 (2021) 153, doi:[10.1007/JHEP11\(2021\)153](https://doi.org/10.1007/JHEP11(2021)153), arXiv:[2107.13021](https://arxiv.org/abs/2107.13021).
- 407 [5] ATLAS Collaboration, “Search for new phenomena in events with an energetic jet and
408 missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”,
409 *Phys. Rev. D* **103** (2021), no. 11, 112006, doi:[10.1103/PhysRevD.103.112006](https://doi.org/10.1103/PhysRevD.103.112006),
410 arXiv:[2102.10874](https://arxiv.org/abs/2102.10874).
- 411 [6] M. Duerr et al., “Hunting the dark Higgs”, *JHEP* **04** (2017) 143,
412 doi:[10.1007/JHEP04\(2017\)143](https://doi.org/10.1007/JHEP04(2017)143), arXiv:[1701.08780](https://arxiv.org/abs/1701.08780).
- 413 [7] N. F. Bell, Y. Cai, and R. K. Leane, “Dark Forces in the Sky: Signals from Z' and the Dark
414 Higgs”, *JCAP* **08** (2016) 001, doi:[10.1088/1475-7516/2016/08/001](https://doi.org/10.1088/1475-7516/2016/08/001),
415 arXiv:[1605.09382](https://arxiv.org/abs/1605.09382).
- 416 [8] F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz, and S. Vogl, “Implications of unitarity and
417 gauge invariance for simplified dark matter models”, *JHEP* **02** (2016) 016,
418 doi:[10.1007/JHEP02\(2016\)016](https://doi.org/10.1007/JHEP02(2016)016), arXiv:[1510.02110](https://arxiv.org/abs/1510.02110).
- 419 [9] N. F. Bell, Y. Cai, and R. K. Leane, “Impact of mass generation for spin-1 mediator
420 simplified models”, *JCAP* **01** (2017) 039, doi:[10.1088/1475-7516/2017/01/039](https://doi.org/10.1088/1475-7516/2017/01/039),
421 arXiv:[1610.03063](https://arxiv.org/abs/1610.03063).
- 422 [10] LHC Dark Matter Working Group, “Recommendations of the LHC dark matter working
423 group: Comparing LHC searches for dark matter mediators in visible and invisible decay
424 channels and calculations of the thermal relic density”, *Phys. Dark Univ.* **26** (2019)
425 100377, doi:[10.1016/j.dark.2019.100377](https://doi.org/10.1016/j.dark.2019.100377), arXiv:[1703.05703](https://arxiv.org/abs/1703.05703).
- 426 [11] M. T. Frandsen et al., “LHC and Tevatron Bounds on the Dark Matter Direct Detection
427 Cross-Section for Vector Mediators”, *JHEP* **07** (2012) 123,
428 doi:[10.1007/JHEP07\(2012\)123](https://doi.org/10.1007/JHEP07(2012)123), arXiv:[1204.3839](https://arxiv.org/abs/1204.3839).
- 429 [12] CMS Collaboration, “A portrait of the Higgs boson by the CMS experiment ten years
430 after the discovery.”, *Nature* **607** (2022), no. 7917, 60,
431 doi:[10.1038/s41586-022-04892-x](https://doi.org/10.1038/s41586-022-04892-x), arXiv:[2207.00043](https://arxiv.org/abs/2207.00043). [Erratum:
432 doi:[10.1038/s41586-023-06164-8](https://doi.org/10.1038/s41586-023-06164-8)].
- 433 [13] ATLAS Collaboration, “Search for dark matter produced in association with a dark Higgs
434 boson decaying into W^+W^- in the one-lepton final state at $\sqrt{s}=13$ TeV using 139 fb^{-1} of

- 435 *pp* collisions recorded with the ATLAS detector”, *JHEP* **07** (2023) 116,
436 doi:10.1007/JHEP07(2023)116, arXiv:2211.07175.
- 437 [14] CMS Collaboration, “Search for dark matter particles in W^+W^- events with transverse
438 momentum imbalance in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **03** (2024) 134,
439 doi:10.1007/JHEP03(2024)134, arXiv:2310.12229.
- 440 [15] ATLAS Collaboration, “Search for dark matter produced in association with a dark Higgs
441 boson in the $b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”,
442 arXiv:2407.10549.
- 443 [16] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004,
444 doi:10.1088/1748-0221/3/08/S08004.
- 445 [17] CMS Collaboration, “Development of the CMS detector for the CERN LHC Run 3”,
446 *JINST* **19** (2024) P05064, doi:10.1088/1748-0221/19/05/P05064.
- 447 [18] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions
448 at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020), no. 10, P10017,
449 doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.
- 450 [19] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020,
451 doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- 452 [20] CMS Collaboration, “Performance of the CMS high-level trigger during LHC Run 2”,
453 *JINST* **19** (2024), no. 11, P11021, doi:10.1088/1748-0221/19/11/P11021,
454 arXiv:2410.17038.
- 455 [21] CMS Collaboration, “Electron and photon reconstruction and identification with the
456 CMS experiment at the CERN LHC”, *JINST* **16** (2021), no. 05, P05014,
457 doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.
- 458 [22] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction
459 with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018) P06015,
460 doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- 461 [23] CMS Collaboration, “Description and performance of track and primary-vertex
462 reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009,
463 doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- 464 [24] CMS Tracker Group Collaboration, “The CMS phase-1 pixel detector upgrade”, *JINST*
465 **16** (2021) P02027, doi:10.1088/1748-0221/16/02/P02027, arXiv:2012.14304.
- 466 [25] CMS Collaboration, “Track impact parameter resolution for the full pseudo rapidity
467 coverage in the 2017 dataset with the CMS phase-1 pixel detector”, CMS Detector
468 Performance Summary CMS-DP-2020-049, 2020.
- 469 [26] CMS Collaboration, “2017 tracking performance plots”, CMS Detector Performance
470 Summary CMS-DP-2017-015, 2017.
- 471 [27] CMS Collaboration, “Technical proposal for the Phase-II upgrade of the Compact Muon
472 Solenoid”, CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.
- 473 [28] CMS Collaboration, “Particle-flow reconstruction and global event description with the
474 CMS detector”, *JINST* **12** (2017), no. 10, P10003,
475 doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

- 476 [29] CMS Collaboration, “Reconstruction and identification of τ lepton decays to hadrons and
477 ν_τ at CMS”, *JINST* **11** (2016) P01019, doi:10.1088/1748-0221/11/01/P01019,
478 arXiv:1510.07488.
- 479 [30] C. Collaboration, “Identification of hadronic tau lepton decays using a deep neural
480 network”, *JINST* **17** (2022) P07023, doi:10.1088/1748-0221/17/07/P07023,
481 arXiv:2201.08458.
- 482 [31] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm”, *JHEP* **04**
483 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- 484 [32] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012)
485 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- 486 [33] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp
487 collisions at 8 TeV”, *JINST* **12** (2017) P02014,
488 doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- 489 [34] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15** (2020) P09018,
490 doi:10.1088/1748-0221/15/09/p09018, arXiv:2003.00503.
- 491 [35] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification”, *JHEP* **10**
492 (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- 493 [36] E. Bols et al., “Jet Flavour Classification Using DeepJet”, *JINST* **15** (2020), no. 12, P12012,
494 doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.
- 495 [37] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet
496 substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029,
497 arXiv:1307.0007.
- 498 [38] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new
499 Higgs search channel at the LHC”, *Phys. Rev. Lett.* **100** (2008) 242001,
500 doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- 501 [39] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft drop”, *JHEP* **05** (2014) 146,
502 doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- 503 [40] CMS Collaboration, “Identification of heavy, energetic, hadronically decaying particles
504 using machine-learning techniques”, *JINST* **15** (2020), no. 06, P06005,
505 doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.
- 506 [41] CMS Collaboration, “Performance of missing transverse momentum reconstruction in
507 proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”, *JINST* **14** (2019)
508 P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.
- 509 [42] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015)
510 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- 511 [43] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from
512 underlying-event measurements”, *Eur. Phys. J. C* **80** (2020), no. 1, 4,
513 doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- 514 [44] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506**
515 (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

- 516 [45] NNPDF Collaboration, “Parton distributions for the LHC run II”, *JHEP* **04** (2015) 040,
517 doi:[10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040), arXiv:[1410.8849](https://arxiv.org/abs/1410.8849).
- 518 [46] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur.
519 Phys. J.* **C77** (2017), no. 10, 663, doi:[10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5),
520 arXiv:[1706.00428](https://arxiv.org/abs/1706.00428).
- 521 [47] J. Alwall et al., “The automated computation of tree-level and next-to-leading order
522 differential cross sections, and their matching to parton shower simulations”, *JHEP* **07**
523 (2014) 079, doi:[10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079), arXiv:[1405.0301](https://arxiv.org/abs/1405.0301).
- 524 [48] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching matrix elements and
525 shower evolution for top-quark production in hadronic collisions”, *JHEP* **01** (2007) 013,
526 doi:[10.1088/1126-6708/2007/01/013](https://doi.org/10.1088/1126-6708/2007/01/013), arXiv:[hep-ph/0611129](https://arxiv.org/abs/hep-ph/0611129).
- 527 [49] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012)
528 061, doi:[10.1007/JHEP12\(2012\)061](https://doi.org/10.1007/JHEP12(2012)061), arXiv:[1209.6215](https://arxiv.org/abs/1209.6215).
- 529 [50] M. Czakon, P. Fiedler, and A. Mitov, “Total top-quark pair-production cross section at
530 hadron colliders through $o(\alpha_s^4)$ ”, *Phys. Rev. Lett.* **110** (2013) 252004,
531 doi:[10.1103/PhysRevLett.110.252004](https://doi.org/10.1103/PhysRevLett.110.252004), arXiv:[1303.6254](https://arxiv.org/abs/1303.6254).
- 532 [51] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with
533 shower in POWHEG: s - and t -channel contributions”, *JHEP* **09** (2009) 111,
534 doi:[10.1088/1126-6708/2009/09/111](https://doi.org/10.1088/1126-6708/2009/09/111), arXiv:[0907.4076](https://arxiv.org/abs/0907.4076). [Erratum:
535 doi:[10.1007/JHEP02\(2010\)011](https://doi.org/10.1007/JHEP02(2010)011)].
- 536 [52] E. Re, “Single-top Wt -channel production matched with parton showers using the
537 POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547,
538 doi:[10.1140/epjc/s10052-011-1547-z](https://doi.org/10.1140/epjc/s10052-011-1547-z), arXiv:[1009.2450](https://arxiv.org/abs/1009.2450).
- 539 [53] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR”,
540 *Comput. Phys. Commun.* **182** (2011) 1034, doi:[10.1016/j.cpc.2010.12.040](https://doi.org/10.1016/j.cpc.2010.12.040),
541 arXiv:[1007.1327](https://arxiv.org/abs/1007.1327).
- 542 [54] P. Kant et al., “HATHOR for single top-quark production: Updated predictions and
543 uncertainty estimates for single top-quark production in hadronic collisions”, *Comput.
544 Phys. Commun.* **191** (2015) 74, doi:[10.1016/j.cpc.2015.02.001](https://doi.org/10.1016/j.cpc.2015.02.001),
545 arXiv:[1406.4403](https://arxiv.org/abs/1406.4403).
- 546 [55] T. Gehrmann et al., “ W^+W^- production at hadron colliders in next to next to leading
547 order QCD”, *Phys. Rev. Lett.* **113** (2014) 212001,
548 doi:[10.1103/PhysRevLett.113.212001](https://doi.org/10.1103/PhysRevLett.113.212001), arXiv:[1408.5243](https://arxiv.org/abs/1408.5243).
- 549 [56] J. M. Campbell and R. K. Ellis, “An update on vector boson pair production at hadron
550 colliders”, *Phys. Rev. D* **60** (1999) 113006, doi:[10.1103/PhysRevD.60.113006](https://doi.org/10.1103/PhysRevD.60.113006),
551 arXiv:[hep-ph/9905386](https://arxiv.org/abs/hep-ph/9905386).
- 552 [57] J. M. Lindert et al., “Precise predictions for V+jets dark matter backgrounds”, *Eur. Phys.
553 J. C* **77** (2017) 829, doi:[10.1140/epjc/s10052-017-5389-1](https://doi.org/10.1140/epjc/s10052-017-5389-1), arXiv:[1705.04664](https://arxiv.org/abs/1705.04664).
- 554 [58] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at
555 $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”, *Eur. Phys. J. C* **81** (2021), no. 9, 800,
556 doi:[10.1140/epjc/s10052-021-09538-2](https://doi.org/10.1140/epjc/s10052-021-09538-2), arXiv:[2104.01927](https://arxiv.org/abs/2104.01927).

- 557 [59] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at
558 $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.
- 559 [60] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at
560 $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.
- 561 [61] CMS Collaboration, “Performance of the CMS missing transverse momentum
562 reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P02006,
563 doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.
- 564 [62] CMS Collaboration, “Search for new physics in final states with an energetic jet or a
565 hadronically decaying W or Z boson and transverse momentum imbalance at
566 $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **97** (2018) 092005, doi:10.1103/PhysRevD.97.092005,
567 arXiv:1712.02345.
- 568 [63] J. S. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multisource
569 Spectra”, in *PHYSTAT 2011*, pp. 115–120. 2011. arXiv:1103.0354.
570 doi:10.5170/CERN-2011-006.115.
- 571 [64] A. L. Read, “Presentation of search results: The CL_s technique”, *J. Phys. G* **28** (2002) 2693,
572 doi:10.1088/0954-3899/28/10/313.
- 573 [65] T. Junk, “Confidence level computation for combining searches with small statistics”,
574 *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2,
575 arXiv:hep-ex/9902006.
- 576 [66] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for
577 likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554,
578 doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum:
579 doi:10.1140/epjc/s10052-013-2501-z].