
CMS Physics Analysis Summary

Contact: cms-pag-conveners-sus@cern.ch

2025/05/16

Search for supersymmetry in hadronic and leptonic final states with highly Lorentz-boosted objects at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search for supersymmetry in final states with highly Lorentz-boosted top quarks, W , Z , Higgs bosons, or leptonic jets is presented. The search is based on proton-proton collision data at a center-of-mass energy of $\sqrt{s} = 13$ TeV, collected by the CMS experiment at CERN LHC during 2016, 2017, and 2018, corresponding to an integrated luminosity of 138 fb^{-1} . Candidates for boosted top quark and W , Z , or Higgs boson decays are identified using jet substructure techniques. In addition, leptonic decays of boosted standard model or supersymmetry particles are explored by identifying boosted leptonic jets. The analysis is performed in channels with zero leptons, an isolated lepton, or a nonisolated lepton. Signal events are discriminated from standard model background events using the razor kinematic variables, which characterize signals with massive particles decaying to visible particles and massive invisible particles as a peak above the smoothly falling background. Standard model backgrounds are estimated by deriving data over simulation correction factors in background-enriched control regions. Data are consistent with standard model expectations. The results are interpreted using several simplified supersymmetry models with pair production of gluinos, top squarks, bottom squarks, and electroweakinos, featuring both R -parity-conserving and R -parity-violating decay chains.

1 Introduction

Supersymmetry (SUSY) [1–9] is a well-motivated extension of the standard model (SM) of particle physics. It provides solutions to several open questions, including the identity of dark matter (DM) [10, 11], the stability of the electroweak scale, and the observed anomaly in the magnetic moment of the muon [12, 13]. SUSY introduces a symmetry between fermions and bosons, predicting a corresponding superpartner for each SM particle. The search for these SUSY particles (sparticles) is a major component of the physics program at the CERN LHC. ATLAS [14] and CMS [15] collaborations have not yet found evidence for SUSY, but have placed increasingly stringent limits on sparticle masses and production cross sections, through a large variety of inclusive or targeted searches exploring the production of colored and noncolored sparticles in varying mass hierarchies and decay structures. These results motivate the continued exploration of scenarios involving heavier sparticles and suppressed cross sections, which span a significant region of the still viable SUSY parameter space.

This note presents an inclusive search for realizations of SUSY with heavy sparticles and large mass splittings, using 138 fb^{-1} of proton-proton collision data collected by the CMS experiment at the CERN LHC during years 2016–2018. In such scenarios, heavy sparticles decay into lighter SUSY or SM particles with substantial Lorentz boosts. If the decay products are themselves unstable, their subsequent decays can result in final state particles that are collimated, that is, close in the η – ϕ plane.

Final states with large mass splittings and boosted decay products arise in a wide variety of SUSY scenarios, particularly those motivated by naturalness, which predicts gluinos and third-generation squarks to have masses of the order of one TeV. Figure 1 shows representative diagrams of gluino, top squark, and electroweak gaugino pair production in models with conserved R -parity. In these examples, the sparticles decay to a much lighter, stable neutralino—the lightest SUSY particle (LSP)—either directly or via two-step decay chains. These decays produce highly boosted heavy SM particles such as W , Z , or Higgs bosons, or top quarks. The final state also includes lighter SM leptons and quarks, and moderate to large momentum imbalance arising from the undetected neutralino LSP.

An alternative class of scenarios involves R -parity violation, where the LSP is unstable and decays to SM particles. In this case, the boosted object is typically a high-momentum R -parity violating LSP that is significantly heavier than its SM decay products. Figure 2 shows two representative examples: gluino pair production with decays to R -parity violating top squarks, and bottom squark pair production with decays to R -parity violating neutralinos. In both cases, the LSP decays into one or more quarks and leptons. As in the R -parity violating case, the final state includes lighter SM particles, but the undetected momentum is reduced compared to the R -parity conserving case, since it originates only from SM neutrinos in the absence of a stable neutral sparticle.

Earlier CMS searches have explored final states containing highly Lorentz-boosted particles, such as top quarks [16, 17], and W , Z , and Higgs bosons [18–20] using 138 fb^{-1} of 13 TeV data. These efforts primarily focused on targeted final states or specific production modes.

The current study adopts a more comprehensive and systematic approach to the exploration of boosted final states. It extends a previous study based on 35.9 fb^{-1} of data [21], which focused exclusively on final states with boosted hadronically decaying W bosons and top quarks. It targets a broad class of SUSY scenarios by incorporating final states with hadronically decaying boosted W , Z , and Higgs bosons, as well as top quarks decaying leptonically via W bosons. In addition, it explores, for the first time in a SUSY context, final states featuring boosted top

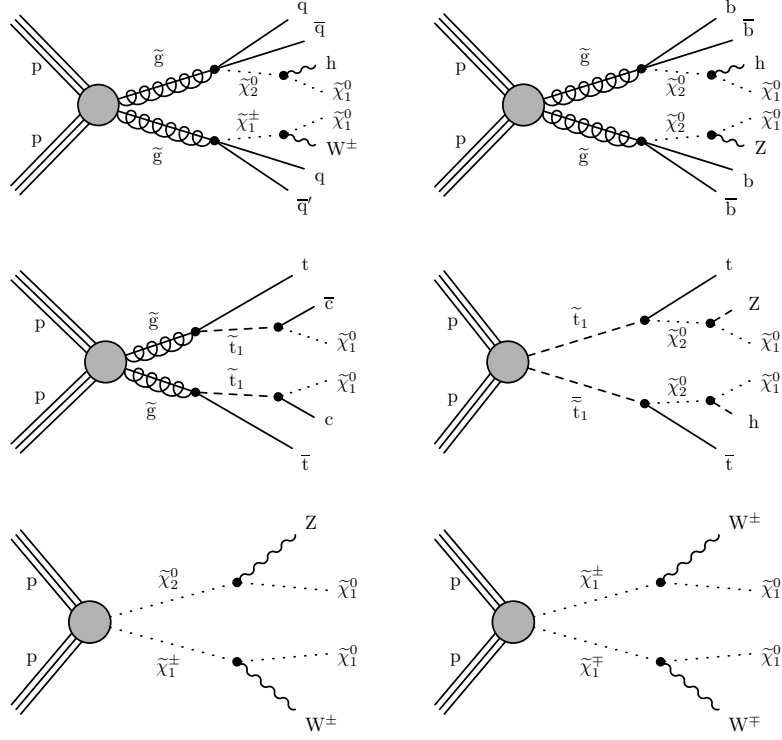


Figure 1: R -parity conserving signal models considered in this analysis: Gluino pair production T5qqqqWH (top left), T5bbbbZH (top right) and T5ttcc (middle left); top squark pair production T6ttZH (middle right); chargino-neutralino production TChiWZ (bottom left) and chargino-chargino production TChiWW (bottom right).

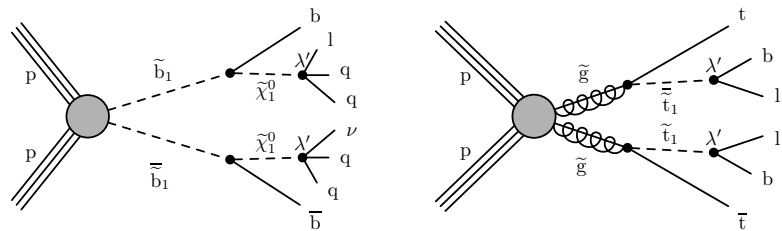


Figure 2: R -parity violating signal models considered in this analysis. Bottom squark pair production decaying to R -parity violating neutralino, R2bbqqlv (left); gluino pair production decaying to R -parity violating top squark, R5ttbl (right).

quarks decaying leptonically, and boosted sparticles (partially) reconstructed as massive jets containing nonisolated leptons, referred to as boosted leptonic jets. The analysis is performed in distinct channels requiring zero leptons, one isolated lepton, or one nonisolated lepton, with the latter accommodating the boosted leptonic jets. Additional selections based on jet and b-tagged jet multiplicities are applied to define a set of exclusive signal regions, each targeting scenarios with different sparticle masses, production or decay modes. Signal events are separated from SM backgrounds using the razor kinematic variables M_R and R^2 [22], which characterize signal-like events as localized peaks above a smoothly falling background. Each signal region is further partitioned into bins defined by the razor variables.

The razor kinematic variables used in this analysis are introduced in Section 2. The data sets and Monte Carlo (MC) simulation samples are described in Section 4. Object definitions and event selection criteria are detailed in Sections 5 and 6, respectively. The background estimation methodology is presented in Section 7, and the systematic uncertainties considered in the analysis are discussed in Section 8. The results and their interpretation are given in Sections 9 and 10.

2 The razor variables

The razor variables M_R and R^2 map the event into a dijet topology [22]. They help to describe a signal coming from the pair production of two heavy particles, each decaying to a massless visible particle and a massive invisible particle, as a peak over exponentially falling SM backgrounds. For this reason, the razor variables are robust discriminators for SUSY signals where the sparticles are pair-produced and subsequently decay to lighter SM particles and the invisible lightest SUSY particle (LSP). For the simple case where the final topology has two visible particles, e.g., jets, the razor variables are defined using the 4-momenta of these two jets as

$$M_R \equiv \sqrt{\left(|\vec{p}_{j1}| + |\vec{p}_{j2}|\right)^2 - \left(p_z^{j1} + p_z^{j2}\right)^2}, \quad (1)$$

$$M_T^R \equiv \sqrt{\frac{p_T^{\text{miss}} \left(p_T^{j1} + p_T^{j2}\right) - \vec{p}_T^{\text{miss}} \cdot \left(\vec{p}_T^{j1} + \vec{p}_T^{j2}\right)}{2}}. \quad (2)$$

Given a global estimator M_R and a transverse estimator M_T^R , the razor dimensionless ratio is defined as

$$R \equiv \frac{M_T^R}{M_R}. \quad (3)$$

If the decay chains are more complicated and there are multiple particles in the final state, we first form two megajets from the final state particles, such that each of the megajets contains the particles coming from one of the heavy pair-produced particles. M_R and R^2 are then computed using the 4-momenta of these two megajets, where the megajet 4-momenta are computed as the sums of the 4-momenta of the jets contributing to each megajet. Of all the possible partitions of the jets into two megajets, we select the combination that minimizes the sum of the invariant masses of the two megajets. This choice will cluster together particles that are traveling in the same direction and has been found to have a good performance. Megajet computation takes as input jets that are reconstructed using the anti- k_T algorithm [23, 24] with a distance parameter of $\Delta R = 0.4$, to be detailed in Section 5. Depending on the selected final state, isolated or

nonisolated electrons or muons, or isolated photons not overlapping with jets are also included in the megajet calculation.

3 The CMS detector

The CMS apparatus [15, 25] is a multipurpose, nearly hermetic detector, designed to trigger on [26–28] and identify electrons, muons, photons, and (charged and neutral) hadrons [29–31]. Its central feature is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are reconstructed using gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. More detailed descriptions of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Refs. [15, 25].

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of 4 μ s [26]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [27].

4 Data and Monte Carlo samples

This analysis uses pp collision data at 13 TeV center-of-mass energy collected during the LHC Run 2 period, encompassing the years 2016, 2017, and 2018. Collected data correspond to integrated luminosities of 36.3, 41.5, and 59.8 fb^{-1} in each year, respectively, leading to a total of 138 fb^{-1} .

The analysis also relies on simulated MC event samples for obtaining signal yields and estimating the SM backgrounds, all of which were generated at $\sqrt{s} = 13$ TeV. For each signal process shown in Figs. 1 and 2, the hard process was generated using MADGRAPH5_aMC@NLO (v2.6.5) [32] at leading order (LO) with up to two additional partons at the matrix element level, while their decays were simulated with PYTHIA (v8.240) [33]. The next-to-next-to-leading order (NNLO) parton distribution function (PDF) set NNPDF3.1 [34] was used. Signal production cross sections are approximated to NNLO plus next-to-next-to-leading-logarithm (NNLL) [35, 36].

Top quark pair production, t -channel single top quark production, and single top quarks produced in association with a W boson are generated with the next-to-leading-order (NLO) generator POWHEG v2.0 [37–40]. The MADGRAPH5_aMC@NLO generator is used in NLO mode for $t\bar{t} + W/Z/H$, $t\bar{t}t\bar{t}$, tZq , and s -channel single top quark production events, while leading-order (LO) mode is used for QCD multijet, $t\bar{t} + \text{diboson}/tW/t + \text{jet}$, Drell–Yan, and $W + \text{jets}$ production. Diboson production is generated with PYTHIA (v8.240). Production cross sections are calculated to the highest available accuracy, generally, NLO or NNLO [32, 41–47].

For modeling fragmentation and parton showering, the generators described above are interfaced to PYTHIA (v8.240). For samples generated at LO (NLO) precision, the MLM [48] (FxFx [49]) prescription is used to match partons from the matrix element calculation to those

from parton showers. The underlying-event tunes CP5 and CP2 [50] are used for the SM background and signal samples, respectively, for all three years. Detector response for the background simulated samples is modeled using the GEANT4 [51] based CMS detector simulation suite, while for the signal samples, the CMS fast simulation package is used [52, 53].

5 Event reconstruction and object identification

The objects used in this analysis are reconstructed starting with a particle-flow (PF) algorithm [54]. The PF algorithm aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all Bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected with the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies. The primary vertex is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in Section 9.4.1 of Ref. [55].

Electrons are reconstructed by associating a track in the inner silicon detectors with a cluster of energy in the ECAL. The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with transverse momentum $p_T \approx 45$ GeV from $Z \rightarrow ee$ decays ranges from 1.6 to 5%. It is generally better in the barrel region than in the endcaps, and also depends on the bremsstrahlung energy emitted by the electron as it traverses the material in front of the ECAL [29, 56]. Electron identification is done by an algorithm based on a boosted decision tree (BDT). The working point where the identification efficiency reaches approximately 0.8 is selected. The analysis uses both isolated and nonisolated electrons. Identification includes (excludes) isolation criteria in the BDT training for isolated (nonisolated) electrons. Isolated electrons are additionally required to pass a relative isolation condition based on the variable $I_{\text{rel}}^{\text{PF}}$, defined as the scalar sum of the p_T of all PF candidates within a cone of size $\Delta R = 0.3$ around the electron, divided by the electron p_T . A requirement of $I_{\text{rel}}^{\text{PF}} < 0.4$ is applied. A subset of these electrons, called isolated select electrons, are required to have a more stringent identification, with efficiency 0.9, and isolation $I_{\text{rel}}^{\text{PF}} < 0.1$. Nonisolated electrons are required to satisfy either $\Delta R(e, \text{jet}) > 0.4$ or $p_T^{\text{rel}} > 15$ GeV, where p_T^{rel} is defined as $|\vec{p}_\ell \times \vec{p}_{\text{jet}}| / |\vec{p}_{\text{jet}}|$. Isolated, isolated select, and nonisolated electrons are required to have $p_T > 5, 10$, and 30 GeV, respectively, and $|\eta| < 2.5$. Electrons within the barrel-endcap transition region, defined by $1.44 < |\eta| < 1.57$, are vetoed. Additionally, transverse and longitudinal impact parameters with respect to the primary vertex should be $|d_{xy}| < 0.05$ cm, and $|d_z| < 0.1$ cm for all electrons.

Muons are reconstructed by associating a track in the inner silicon detectors with a track in the muon system. They are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with p_T up to 100 GeV, of 1% in the barrel and 3% in the endcaps. The p_T resolution in the barrel is better than 7% for muons with p_T up to 1 TeV [30]. Muon identification is done by applying a set of quality criteria on the tracks reconstructed in

the tracker and the muon detectors. The loose working point is selected, which has an identification efficiency of 0.90. Isolated and nonisolated muons are defined with the same conditions as for electrons, with the exception that the cone defining relative isolation for muons having a radius of $\Delta R = 0.4$. Similarly, isolated select muons are defined, with more stringent, medium identification, with efficiency 0.95, and isolation $I_{\text{rel}}^{\text{PF}} < 0.2$. Isolated, isolated select, and nonisolated muons are required to have $p_T > 5, 10$, and 30 GeV, respectively, and $|\eta| < 2.4$. Additionally, transverse and longitudinal impact parameters with respect to the primary vertex should be $|d_{xy}| < 0.2, 0.05, 0.2$ cm, and $|d_z| < 0.5, 0.1, 0.5$ cm for isolated, isolated select, and nonisolated muons, respectively.

To suppress backgrounds from τ lepton decays, charged particle tracks not identified as isolated electrons or muons, including PF electrons and muons, are subjected to a track isolation requirement. A track is considered isolated if the scalar p_T sum of all other charged tracks within a cone of radius 0.3 around its direction, divided by the track p_T , is less than 0.2 for PF electrons or muons, and less than 0.1 otherwise. Isolated tracks are required to satisfy $|\eta| < 2.4$.

Photons, used exclusively for background estimation, are identified using a cut-based selection with the medium working point, corresponding to an average identification efficiency of 80%. Photons are required to have $p_T > 80$ GeV, $|\eta| < 2.5$, and must satisfy the electron veto and pixel conversion rejection criteria.

Jets are reconstructed from the PF particles using the infrared and collinear safe anti- k_T algorithm [23, 24] with a distance parameter of $\Delta R = 0.4$. 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. Additional pp interactions within the same or nearby bunch crossings can contribute additional tracks and calorimetric energy depositions, increasing the apparent jet momentum. To mitigate this effect, tracks identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions [57]. Jet energy corrections are derived from simulation studies so that the average measured energy of jets becomes identical to that of particle level jets. In situ measurements of the momentum balance in dijet, photon + jet, Z + jet, and multijet events are used to determine any residual differences between the jet energy scale in data and in simulation, and appropriate corrections are made [58]. Additional selection criteria are applied to each jet to remove jets potentially dominated by instrumental effects or reconstruction failures [57]. A jet-lepton cleaning that excludes jets overlapping with selected leptons within a cone of $\Delta R = 0.4$ is applied. A tight identification is applied by imposing criteria on variables related to energy fractions and multiplicities of various PF candidate types clustered in a jet, in order to distinguish physical jets from those arising from calorimetry noise. Jets are selected to have $p_T > 30$ GeV and $|\eta| < 2.4$.

Jets originating from the hadronization of bottom quarks, denoted as b-tagged jets are identified using the DeepJet [59–61] algorithm. For jets passing the above jet criteria, the medium working point having an average b-tagging efficiency of ≈ 0.7 and a gluon and light flavor jet misidentification rate of 0.01, is used for selecting events with b-tagged jets. The looser working point, with efficiency of ≈ 0.85 , and misidentification rate of 0.1, is used for vetoing events with b-tagged jets.

The missing transverse momentum vector $\vec{p}_T^{\text{miss}}$ is defined as the projection onto the plane perpendicular to the beam axis of the negative vector sum of the momenta of all reconstructed particle-flow objects in an event. Its magnitude is referred to as p_T^{miss} [62]. The $\vec{p}_T^{\text{miss}}$ vector is modified to account for corrections to the energy scale of the reconstructed jets in the event. The pileup per particle identification (PUPPI) algorithm [63] is applied to reduce the pileup

dependence of the $\vec{p}_T^{\text{miss}}$ observable.

5.1 Boosted objects

The analysis makes use of both hadronically and leptonically decaying boosted objects, which are reconstructed using the anti- k_T algorithm with a size parameter of 0.8 (AK8). The hadronic category includes boosted W , Z , and Higgs bosons, as well as top quarks. These objects are identified using the ParticleNet tagger [64], a deep neural network trained to distinguish between boosted hadronic decays of heavy SM particles and jets originating from QCD multijet processes. Specific requirements on the jet mass and transverse momentum are applied to ensure consistency with the kinematics of highly Lorentz-boosted decays. The jet mass is computed using the soft-drop grooming algorithm [65], which reduces contamination from soft, wide-angle radiation and pileup. For boosted W bosons and top quarks, a soft-drop mass window of 50–220 GeV is used, with a narrower window of 70–105 GeV for Z bosons and 90–140 GeV for Higgs bosons. Transverse momentum thresholds are chosen to ensure that the decay products of the parent particle are fully contained within the jet cone ($R = 0.8$), with minimum p_T values ranging from 200 to 300 GeV, depending on the object. Dedicated discriminants provided by the ParticleNet tagger are used to distinguish W , Z , and H bosons, as well as top quarks, from light-flavor jets, and selection criteria based on these discriminants are applied to identify the corresponding boosted objects.

A novel feature of this analysis is the inclusion of leptonic boosted objects. These are defined as AK8 jets with $p_T > 300$ GeV and a soft-drop mass greater than 50 GeV that contain at least one nonisolated lepton within a cone of $\Delta R < 0.8$ around the jet axis. These objects target leptonic decays of boosted top quarks or hypothetical supersymmetric particles, such as R -parity-violating (RPV) neutralinos. No dedicated multivariate tagger is used for these objects; their definition relies solely on geometric and kinematic association with a nearby lepton.

In addition to explicitly tagged categories, the analysis also makes use of a more inclusive set of mass-tagged AK8 jets, defined by a requirement of $p_T > 200$ GeV and a soft-drop mass $m > 50$ GeV. These are primarily used in the definition of control and validation regions.

Because the ParticleNet taggers are trained against QCD jets and are not mutually exclusive, a single AK8 jet may satisfy the tagging criteria for more than one boosted object. To resolve such overlaps and ensure a unique classification, a priority ordering is applied: i) leptonic jet, ii) hadronic Higgs boson, iii) hadronic top quark, and iv) hadronic W or Z boson. If a jet satisfies multiple tagging criteria, it is assigned the first applicable label in this list. In particular, if a jet is tagged both as a hadronic top quark and as a Higgs boson, it is classified as a Higgs boson. This prioritization reflects the focus of the analysis on boosted leptonic jets and Higgs bosons. The former are incorporated for the first time in a SUSY search using nonisolated leptons, enabling sensitivity to scenarios with high-mass RPV neutralinos or lepton-rich topologies. The latter extend beyond earlier searches limited to specific final states such as $HH + p_T^{\text{miss}}$, allowing a more inclusive treatment of boosted H bosons that enhances coverage of a broader set of SUSY models with varying decay structures.

6 Event preselection and signal regions

This analysis involves a large set of event selections categorized into signal regions (SRs) that enhance sensitivity to signals, control regions (CRs) to estimate the SM backgrounds, and validation regions (VRs) to validate background estimation. Events in the SRs, as well as in all but one CRs, and all VRs are selected using a combination of triggers based on the scalar sum of jet

transverse momenta (denoted as H_T), the p_T of AK8 jets, and p_T^{miss} . The H_T components have thresholds ranging from 900 to 1050 GeV, while the AK8 jet triggers require $p_T > 450$ GeV, and the p_T^{miss} triggers have thresholds between 75 and 100 GeV. These triggers are not fully efficient across the entire preselection phase space. To account for this, their combined efficiency is measured from data in two dimensions, in bins of p_T^{miss} versus H_T , using samples selected with orthogonal trigger paths requiring at least one electron or muon. The resulting efficiency maps are applied as event-by-event weights to simulated events to correct for the trigger turn-on behavior.

Table 1 lists the baseline preselection applied before all SR, CR, and VR selections, as well as the additional selections used to define the fully hadronic (had), isolated leptonic (lep), and leptonic jet (lepJet) SR categories.

Table 1: Event preselection and SR categories

Preselection			
good primary vertices	> 0		
number of AK8 jets	≥ 1		
number of AK4 jets	≥ 2		
M_R	$> 800 \text{ GeV}$		
R^2	> 0.08		
SR categories:	had	lep	lepJet
number of iso select leptons $\ell_{\text{isel}} : e, \mu$	$= 0$	$= 1$	$-$
number of iso tracks	$= 0$	$= 0$	$= 0$
number of leptonic jets	$= 0$	$= 0$	≥ 1
$\Delta\phi_R \equiv \Delta\phi_{\text{megajets}}$	< 2.8	$-$	$-$
m_T	$-$	$> 120 \text{ GeV}$	$> 140 \text{ GeV}$
$\Delta\phi_{J,W} \equiv \Delta\phi(\vec{j}_{\text{AK8}}^{\text{lead}}, \vec{\ell} + \vec{p}_T^{\text{miss}})$	$-$	$-$	> 0.8

The SRs are defined by the presence of at least one tagged boosted object, selected from five categories: hadronic W, Z, or Higgs bosons, hadronic top quarks, and leptonic jets. The full set of SR selections are listed in Table 2. Events are further classified according to the multiplicities of isolated leptons, AK4 jets, b -tagged jets, and each type of tagged boosted object. Each SR is designed to target specific SUSY production and decay scenarios, optimizing sensitivity across a wide range of signal models.

In the hadronic SRs, background contributions from SM multijet production are suppressed by requiring the azimuthal angle between the two megajets used to compute the razor variables to be $\Delta\phi_R \equiv \Delta\phi_{\text{megajets}} < 2.8$. In the isolated leptonic SRs, to enhance separation from backgrounds involving leptonically decaying top quarks or W bosons, the transverse mass computed using the lepton and $\vec{p}_T^{\text{miss}}$ is required to be $m_T > 120 \text{ GeV}$. In the nonisolated leptonic regions, backgrounds from W boson decays and multijet production are further reduced by requiring $m_T > 140 \text{ GeV}$ and $\Delta\phi_{J,W} \equiv \Delta\phi(\vec{j}_{\text{AK8}}^{\text{lead}}, \vec{\ell} + \vec{p}_T^{\text{miss}}) > 0.8$, respectively. Here, $\Delta\phi(\vec{j}_{\text{AK8}}^{\text{lead}}, \vec{\ell} + \vec{p}_T^{\text{miss}})$ is defined as the azimuthal separation between the leading boosted jet and the reconstructed W boson candidate, which is defined as the vector sum of the leading non-isolated lepton and $\vec{p}_T^{\text{miss}}$. All SRs are defined to be mutually exclusive. Each SR is further subdivided into bins in a transformed product of the razor variables, $(M_R - 800) \times (R^2 - 0.08)$. Bin boundaries for each SR are listed in Table 3.

Table 2: SRs and their selection criteria

	SR	category	ℓ_{isel}	jets	b	V	H	top	lepJet
1	had 1hTop	had	0	≥ 5	–	–	0	1	0
2	had 2hTop	had	0	≥ 2	–	–	0	≥ 2	0
3	had V b 45j	had	0	4–5	≥ 1	≥ 1	0	0	0
4	had V b 6j	had	0	≥ 6	≥ 1	≥ 1	0	0	0
5	had 1V 0b 34j	had	0	3–4	0	1	0	0	0
6	had 1V 0b 5j	had	0	≥ 5	0	1	0	0	0
7	had 2V 0b 24j	had	0	2–4	0	≥ 2	0	0	0
8	had 2V 0b 5j	had	0	≥ 5	0	≥ 2	0	0	0
9	had H b 45j	had	0	4–5	≥ 1	–	≥ 1	–	0
10	had H b 6j	had	0	≥ 6	≥ 1	0	≥ 1	–	0
11	had HV b 6j	had	0	≥ 6	≥ 1	1	1	–	0
12	had H 0b 34j	had	0	3–4	0	0	≥ 1	–	0
13	had H 0b 5j	had	0	≥ 5	0	0	≥ 1	–	0
14	had HV 0b 24j	had	0	2–4	0	≥ 1	≥ 1	–	0
15	had HV 0b 5j	had	0	≥ 5	0	≥ 1	≥ 1	–	0
16	lep 1hTop	lep	1	≥ 3	–	–	0	1	0
17	lep V b	lep	1	≥ 4	≥ 1	≥ 1	0	0	0
18	lep V 0b	lep	1	≥ 2	0	≥ 1	0	0	0
19	lep H b 3j	lep	1	≥ 3	≥ 1	–	≥ 1	–	0
20	lep H 0b 2j	lep	1	≥ 2	0	–	≥ 1	–	0
21	lepJet hTop	nisolep	–	≥ 3	–	–	–	≥ 1	≥ 1
22	lepJet 0V 24j	nisolep	–	2–4	–	0	–	0	≥ 1
23	lepJet 0V 5j	nisolep	–	≥ 5	–	0	–	0	≥ 1
24	lepJet 1V 24j	nisolep	–	2–4	–	1	–	0	≥ 1
25	lepJet 1V 5j	nisolep	–	≥ 5	–	1	–	0	≥ 1

Table 3: SRs and their binning

	SR	Binning in $(M^R - 800) \times (R^2 - 0.08)$ GeV
1	had 1hTop	30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 3000
2	had 2hTop	30, 40, 50, 75, 100, 150, 200, 250, 3000
3	had V b 45j	30, 40, 50, 75, 100, 150, 250, 3000
4	had V b 6j	30, 40, 50, 75, 100, 150, 200, 300, 3000
5	had 1V 0b 34j	30, 40, 50, 75, 100, 150, 200, 250, 400, 3000
6	had 1V 0b 5j	30, 40, 50, 75, 100, 150, 200, 250, 3000
7	had 2V 0b 24j	30, 75, 3000
8	had 2V 0b 5j	30, 50, 3000
9	had H b 45j	30, 40, 50, 75, 100, 150, 200, 250, 3000
10	had H b 6j	30, 40, 50, 75, 100, 150, 200, 3000
11	had HV b 6j	30, 3000
12	had H 0b 34j	30, 40, 50, 75, 100, 150, 200, 3000
13	had H 0b 5j	30, 40, 50, 75, 150, 3000
14	had HV 0b 24j	30, 3000
15	had HV 0b 5j	30, 3000
16	lep 1hTop	30, 40, 50, 75, 100, 150, 200, 250, 350, 3000
17	lep V b	30, 40, 50, 75, 100, 150, 200, 3000
18	lep V 0b	30, 40, 50, 75, 100, 3000
19	lep H b 3j	30, 40, 50, 75, 100, 3000
20	lep H 0b 2j	30, 50, 75, 100, 3000
21	lepJet hTop	30, 40, 50, 75, 100, 150, 200, 250, 3000
22	lepJet 0V 24j	30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 450, 3000
23	lepJet 0V 5j	30, 40, 50, 75, 100, 150, 200, 250, 350, 3000
24	lepJet 1V 24j	30, 40, 50, 75, 100, 3000
25	lepJet 1V 5j	30, 40, 50, 75, 100 150, 3000

7 Background estimation

The SR selections described above significantly suppress the SM backgrounds. However, a nonnegligible contribution remains from several irreducible sources. These backgrounds are classified into the following categories:

- **Multijet:** QCD production of dijet or multijet final states can enter the SRs as a result of jet mismeasurement, which leads to a spurious imbalance of transverse momentum and results in larger values of R^2 . Boosted object tags in these events primarily arise from mistagging light-flavor jets. In rare cases, genuine hadronically decaying W bosons or top quarks can be produced and subsequently tagged. Leptonic boosted object candidates may also appear in such events, originating from non-isolated leptons produced in semileptonic decays of heavy-flavor hadrons or from leptons within hadronic showers.
- **Lost lepton:** In processes such as $t\bar{t} + \text{jets}$, single top production, or $W + \text{jets}$, where a W boson decays leptonically, the lepton may fail to be reconstructed or identified. These events typically feature genuine $\vec{p}_T^{\text{miss}}$ from the neutrino as well as from the undetected lepton. Boosted objects in this category can include genuine hadronic W boson or top quark decays, or misidentified jets mimicking these objects. For instance, in $W(\rightarrow \ell\nu) + \text{jets}$ events, a light-flavor jet may be mistagged as a hadronic boosted object. Additionally, genuine leptonic boosted jets may be present if a high- p_T lepton is merged with nearby hadronic activity, as in boosted top quark decays.
- **$Z(\rightarrow \nu\nu) + \text{jets}$:** In events where a Z boson decays invisibly to neutrinos and is produced in association with jets, the final state can feature large genuine $\vec{p}_T^{\text{miss}}$ and resemble SUSY signatures. This process is particularly relevant in hadronic SRs, especially those with zero b -tagged jets. Boosted hadronic objects in these events originate exclusively from mistagged light-flavor jets. Leptonic boosted jets may also be reconstructed in such events due to nonprompt leptons arising from associated heavy-flavor jet decays.

In the hadronic SRs, the dominant contribution from $t\bar{t} + \text{jets}$ arises from semileptonic decays. Here the top quark decaying to hadronic final state is tagged as a boosted top quark, and the top quark decaying to a leptonic final state yields a lost lepton. Fully hadronic decays contribute minimally in most regions. For single top and $W + \text{jets}$ processes, the contributions from hadronic and leptonic decay modes are generally comparable. In the isolated leptonic SRs, $t\bar{t} + \text{jets}$ events arise predominantly from a balanced mixture of semileptonic and dileptonic decays, while hadronic decays contribute negligibly. In contrast, for single top processes such as tW , and for $W + \text{jets}$ backgrounds, leptonic decays are the dominant source. In the leptonic jet SRs, $t\bar{t} + \text{jets}$ events are primarily from dileptonic decays, with hadronic decays again being negligible. As in the isolated lepton case, single top and $W + \text{jets}$ backgrounds are dominated by leptonic decays. This understanding of background composition is used in the design of CRs that accurately model the processes contributing to each SR.

7.1 Overview of the estimation strategy

The analysis defines 25 SRs with diverse final-state topologies. Relative contributions of irreducible SM backgrounds vary across these regions depending on the lepton category, jet multiplicities, and boosted object content. Nevertheless, a mostly unified background estimation strategy is feasible, and employed across all SRs.

The dominant backgrounds from top quark processes ($t\bar{t} + \text{jets}$ and single top production),

$W + \text{jets}$, QCD multijet, and $Z(\nu\bar{\nu}) + \text{jets}$ are estimated using an MC-assisted, data-driven approach. For each background process, CRs are defined such that the background of interest is kinematically similar to the SR and is the dominant contribution. In each CR, correction factors (CFs) are computed as the ratio of the observed data to the MC prediction in one or more variables of interest. These CFs are then applied to the corresponding background prediction in the SRs.

To mitigate statistical limitations, CFs are derived iteratively in bins of a single variable at a time, rather than in multidimensional bins spanning multiple variables. At each iteration, the CF from the previous step is applied to the MC prediction in the CR, and the next CF is computed versus a new variable. For QCD multijet, top quark, and $W + \text{jets}$ backgrounds, CFs are first determined as a function of $M_R \times R^2$, followed by the number of AK4 jets. This is done separately for events with at least one AK8 jet having a soft-drop mass greater than 50 GeV. For the special case of $Z(\nu\bar{\nu}) + \text{jets}$, estimated using a $\gamma + \text{jets}$ CR, a third iteration is introduced. After applying CFs versus $M_R \times R^2$ and AK4 jet multiplicity, a final CF is derived as a function of the number of AK8 jets with soft-drop mass above 50 GeV. For $Z(\nu\bar{\nu}) + \text{jets}$, CFs are not split by the number of mass-tagged jets, as the template fit method used requires high event yields in each CF bin to ensure statistical robustness.

Potential correlations introduced by the order of the iterative CF computation are assessed by reversing the order of the $M_R \times R^2$ and AK4 jet multiplicity steps. Any discrepancies observed in the resulting background predictions are propagated as a systematic uncertainty. For minor backgrounds contributing at the percent level or below in the SRs, the predictions are taken directly from simulation.

While the general background estimation strategy is unified, slight differences are applied across the three SR categories, primarily arising from the varying definitions of CRs from which CFs are derived. These differences are outlined below:

1. Hadronic SRs:

- Multijet, top quark, and $W + \text{jets}$ backgrounds: CRs enriched in each of these processes are defined by inverting the criteria for various SR selections. CFs are computed simultaneously across the three CRs.
- $Z(\nu\bar{\nu}) + \text{jets}$: CFs are derived using CRs enriched in $\gamma + \text{jets}$ and Drell-Yan events. Cross checks are performed using single lepton CRs enriched in $W + \text{jets}$ and $t\bar{t} + \text{jets}$ processes.
- Other backgrounds: Estimated using MC simulation.

2. Isolated leptonic SRs: For all background processes, the same CFs as those derived for the hadronic SRs are used.

3. Nonisolated leptonic SRs with boosted leptonic jets:

- Top quark and $W + \text{jets}$ backgrounds: CFs are computed using CRs defined by inverting the m_T and $\Delta\phi_{J,W}$ requirements in a modified version of the lepJet preselection, in which the leptonic jet requirement is replaced by the presence of at least one nonisolated lepton.
- Multijet background: CFs from the hadronic SRs are reused.
- Other backgrounds: Estimated using MC simulation.

Table 4 summarizes the CRs along with the associated background processes, selection criteria, and the corresponding SR categories to which they apply. CRs are constructed by inverting

one or more selection criteria that commonly define the corresponding SR category, thereby enriching the contribution of a specific background while maintaining similarity to the SR kinematics. In many cases, CRs are defined separately for events with exactly one AK8 jet and for those with two or more AK8 jets having soft-drop mass greater than 50 GeV. In other cases, this classification is not performed to preserve statistical precision, particularly in regions with limited event yields. All CRs are defined to be orthogonal to the SRs by construction.

Table 4: CRs used for background estimation. The region names W4Z and T4Z are abbreviations for WforZ and TforZ. b_l , b_m : b-quark jets identified with loose and medium working points. CRs 9–14 have additional selection criteria, corresponding to the ', ', and ''' signs next to the CR name: ': Photon added to $\vec{p}_T^{\text{miss}}$, ': Leptons added to $\vec{p}_T^{\text{miss}}$ and $|\text{mass}(\ell\ell) - \text{mass}(Z)| < 10 \text{ GeV}$, ''': Lepton added to $\vec{p}_T^{\text{miss}}$.

	Selection												Used for SR		
	CR	BG	AK8 _{m50}	ℓ_{niso}	b_l	b_m	ℓ_{iso}	ℓ_{isel}	γ	$\Delta\phi_R$	$\Delta\phi_{J,W}$	m_T	Had	lep	lepJet
1	Q1	QCD	1	–	0	–	0	–	–	≥ 2.8	–	–	✓	✓	✓
2	Q2	QCD	≥ 2	–	0	–	0	–	–	≥ 2.8	–	–	✓	✓	✓
3	T1	Top	1	–	–	≥ 1	–	1	0	< 2.8	–	< 120	✓	✓	–
4	T2	Top	≥ 2	–	–	≥ 1	–	1	0	< 2.8	–	< 120	✓	✓	–
5	TnI	Top	≥ 1	≥ 1	–	≥ 1	–	–	0	–	< 0.8	< 140	–	–	✓
6	W1	W	1	–	0	–	–	1	0	< 2.8	–	30–100	✓	✓	–
7	W2	W	≥ 2	–	0	–	–	1	0	< 2.8	–	30–100	✓	✓	–
8	WnI	W	≥ 2	≥ 1	0	–	–	–	0	–	< 0.8	< 140	–	–	✓
9	G'	$Z\nu\bar{\nu}$	≥ 1	–	–	–	0	–	≥ 1	< 2.8	–	–	✓	✓	–
10	DY'''	$Z\nu\bar{\nu}$	≥ 1	–	0	–	–	2,OS	0	< 2.8	–	–	✓	✓	–
11	W4Z1'''	$Z\nu\bar{\nu}$	1	–	0	–	1	–	0	< 2.8	–	30–100	✓	✓	–
12	W4Z2'''	$Z\nu\bar{\nu}$	≥ 2	–	0	–	1	–	0	< 2.8	–	30–100	✓	✓	–
13	T4Z1'''	$Z\nu\bar{\nu}$	1	–	–	≥ 1	1	–	0	< 2.8	–	< 120	✓	✓	–
14	T4Z2'''	$Z\nu\bar{\nu}$	≥ 2	–	–	≥ 1	1	–	0	< 2.8	–	< 120	✓	✓	–

Except for the CF for the $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background, which is derived from a $\gamma + \text{jets}$ CR, all CFs are determined simultaneously for groups of two or three background processes using the corresponding number of CRs. For a given set of N_{BG} major backgrounds, N_{BG} CRs are defined, each selected to be dominated by one of the background processes under consideration. In each CR i and bin j , the observed yield in data is modeled as

$$\text{Data}^{\text{CR}^{ij}} = \sum_{k=1}^{N_{\text{BG}}} \text{CF}_{k,j} \cdot \text{MC}_{k,j}^{\text{CR}^{ij}} + \text{MC}_{\text{other BGs}}^{\text{CR}^{ij}} \quad (4)$$

where $\text{MC}_{k,j}^{\text{CR}^{ij}}$ is the predicted contribution from background k , and $\text{MC}_{\text{other BGs}}^{\text{CR}^{ij}}$ includes all other minor background components. The CFs $\text{CF}_{k,j}$ are extracted by performing a binned likelihood fit to the data in all CRs, independently for each bin j . The fit includes Poisson likelihood terms and accounts for statistical uncertainties in both the data and simulated yields, allowing for correlations between the fitted processes. This strategy enables the simultaneous estimation of multiple background components while controlling the statistical uncertainty in each bin.

7.2 Estimation of the multijet, top quark, and W + jets backgrounds

CFs used to adjust the MC predictions for the QCD multijet background in all SRs, and for the top quark and W + jets backgrounds in the hadronic and isolated leptonic SRs, are derived using the CRs Q1, Q2; T1, T2; and W1, W2, respectively, as listed in Table 4. These CRs are designed to be enriched in multijet, $t\bar{t}$ and single top, and W + jets events, respectively. Two sets of CFs, one for events with exactly one AK8 jet with soft-drop mass greater than 50 GeV,

and another for events with two or more such AK8 jets, are obtained via simultaneous fits to the CR groups (Q1, T1, W1) and (Q2, T2, W2), respectively. For each case, CFs are computed first as a function of $M_R \times R^2$, which are then applied to the corresponding MC background predictions. In the second iteration, new CFs are computed as a function of the number of AK4 jets, and applied on top of the already $M_R \times R^2$ -corrected distributions. Validation of the CFs and background modeling in these CRs is presented in Section 7.5.

7.3 $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background estimation

A direct CR selection isolating the $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background is not feasible. Instead, processes with similar kinematic properties are used to derive CFs for this background. The estimation follows a well-established method that relies on $\gamma + \text{jets}$ - and DY-enriched CRs to compute the central CFs applied to the SRs. In addition, $W(\rightarrow \ell\nu) + \text{jets}$ -enriched single-lepton CRs are used to derive an independent set of CFs, which serve as cross-checks and to evaluate associated systematic uncertainties. The estimate based on the γ and DY CRs is used as the central prediction. The difference between this estimate and the one derived from the $W + \text{jets}$ -based CRs (W4Z/T4Z) is taken as a systematic uncertainty.

7.3.1 Deriving $\text{CF}_{Z\nu\bar{\nu}}$ from $\gamma + \text{jets}$ and dilepton CRs

The main estimate for the CF to be applied to $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ MC events is given by

$$\text{CF}_{Z\nu\bar{\nu}} = \text{CF}_\gamma \cdot R_{Z\ell\ell/\gamma}^{\text{data/MC}} = \text{CF}_\gamma \cdot \frac{N_{Z\ell\ell}^{\text{DY,data}} / N_{Z\ell\ell}^{\text{DY,MC}}}{N_\gamma^{\text{G,data}} / N_\gamma^{\text{G,MC}}}, \quad (5)$$

where the first term, CF_γ , is derived from a photon-enriched CR, and the second term is a double ratio $R_{Z\ell\ell/\gamma}^{\text{data/MC}}$ computed using a DY-enriched CR and the same photon-enriched CR. The double ratio acts as a constant correction that accounts for differences in the cross sections between the $\gamma + \text{jets}$ and $Z(\rightarrow \ell\ell) + \text{jets}$ processes. The quantity CF_γ , whose derivation is detailed below, is computed iteratively as a function of $M_R \times R^2$, the number of AK4 jets, and the number of mass-tagged AK8 jets. The full event-level weight is given by

$$\text{CF}_\gamma = \text{CF}_\gamma^{M_R \times R^2} \cdot \text{CF}_\gamma^{N_{\text{AK4}}} \cdot \text{CF}_\gamma^{N_{\text{AK8}(m>50)}} \quad (6)$$

Selection for the photon-enriched CR "G" is listed in Table 4. The photon is vectorially added to $\vec{p}_T^{\text{miss}}$ in order to mimic the kinematic properties of $\nu\bar{\nu}$ coming from the $Z \rightarrow \nu\bar{\nu}$ decays. Furthermore, jets are removed when they overlap with the photon within an angular separation of $\Delta R < 0.4$ and have a similar p_T to that of the photon, i.e. when the photon p_T to jet p_T ratio is between 0.5 and 2. This procedure aims to avoid double counting photons misidentified as jets. The modified $\vec{p}_T^{\text{miss}}$ and jets are used as input to M_R and R^2 calculation. The main process that contributes to the G region apart from $\gamma + \text{jets}$ is QCD multijet, where a jet is misidentified as a photon. The $\gamma + \text{jets}$ events contain both direct photons and fragmentation photons. Instead of relying on simulation for the subtraction of the multijet background, a data-driven measurement of the photon purity is performed. Data in the G regions are scaled with an estimated photon purity P_γ to exclude the multijet background with misidentified photons, and subsequently with the fraction of fragmentation photons, f , to obtain a pure direct $\gamma + \text{jets}$ sample. Consequently, CF_γ can be expressed as

$$\text{CF}_\gamma = \frac{N_\gamma^{\text{G,data}} \cdot P_\gamma \cdot f}{N_\gamma^{\text{G,MC}}}. \quad (7)$$

The QCD multijet background contamination, and by extension the photon purity, in the G region is estimated using a template fit to the distribution of the photon charged isolation variable. These fits are performed in bins of the variable for which the CF is being derived. Specifically, fits are carried out first as a function of $M_R \times R^2$, then as a function of AK4 jet multiplicity, and finally as a function of mass-tagged AK8 jet multiplicity. Separate fits are performed for the barrel and endcap regions, within $|\eta| < 2.5$. Templates for the QCD multijet process are derived from data in each bin by inverting the $\sigma_{i\eta i\eta}$ selection (requiring $\sigma_{i\eta i\eta} > 0.01015$ for barrel photons and > 0.0272 for endcap photons), while applying all other photon selection criteria. Apart from this modified photon definition, the event selection remains identical to that of the G region. Templates for prompt photons are obtained from simulated γ +jets events. Subsequently, the purity in each bin is computed as $P_\gamma = N_{\text{prompt}} / (N_{\text{prompt}} + N_{\text{misidentified}})$.

The fragmentation photon fraction is directly obtained from the γ +jets MC. Direct prompt photons are defined as prompt photons with an angular separation $\Delta R > 0.4$ with respect to any outgoing quarks or gluons from the hardest subprocess (identified with a PYTHIA status code of 23), and fragmentation prompt photons are all prompt photons that are not direct. The direct photon distribution thus obtained from the G region sample is then used for the z_{inv} estimation. We apply a 10% systematic uncertainty on the fragmentation photon fraction values.

We next use a $Z \rightarrow \ell^+ \ell^-$ CR called "DY" to correct CF_γ . The DY CR, also defined in Table 4, requires two oppositely charged same flavor leptons, whose invariant mass lies within 10 GeV of the Z boson mass. For the e^+e^- case, any extra veto muons are vetoed, while for the $\mu^+\mu^-$ case, any extra electrons that match the veto electron criteria are vetoed. Momentum of the dilepton system is added to $\vec{p}_T^{\text{miss}}$, that, in turn, is used for a modified $R_{\ell\ell}^2$ calculation. The number of loosely b-tagged jets is required to be zero in order to eliminate top quark backgrounds. The integrated number of events in the DY and G regions are used for calculating the double ratio $R_{Z\ell\ell/\gamma}^{\text{data}/\text{MC}}$ as defined in Eq. (6). The resulting double ratios and uncertainties are 0.82 ± 0.17 , 1.29 ± 0.22 , and 1.03 ± 0.17 for the years 2016, 2017, and 2018, respectively. The data-MC discrepancies are reflected into the double ratio as a systematic uncertainty, calculated as the average of the deviations in the $M_R \times R^2$ bins.

7.3.2 Alternative $CF_{Z\nu\bar{\nu}}$ derivation from $W(\rightarrow \ell\nu)$ + jets enriched CRs

To validate the estimates from the G and DY regions, we perform an alternative estimation using a set of CRs enriched in W + jets and semileptonic $t\bar{t}$ + jets, referred to as W4Z and T4Z (abbreviations for W-for-Z and T-for-Z). These CRs, defined separately for cases with exactly one or at least one mass-tagged AK8 jet, are summarized in Table 4. They require exactly one lepton, and either the absence or presence of a b-tagged jet, respectively. The lepton momentum is vectorially added to $\vec{p}_T^{\text{miss}}$ to mimic the contribution of the second neutrino from $Z \rightarrow \nu\bar{\nu}$ decays. The modified $\vec{p}_T^{\text{miss}}$ is then used to recompute the R^2 variable. These selections are identical with those in the W and T regions, with the exception of the modifications introduced by the lepton veto and the recalculated R^2 . We perform a simultaneous fits to the W4Z1, T4Z1, and W4Z2, T4Z2 region pairs. The CF correcting the W + jets process is $CF_{Z\nu\bar{\nu},W}$.

7.4 Background estimation for the leptonic jet regions

SRs in the nonisolated lepton category are dominated by top quark backgrounds, followed by multijet and W + jets processes. The CF for the multijet background is taken from the estimation described in Section 7.2. In contrast, CFs for the top quark and W + jets backgrounds in the nonisolated lepton category are derived using the simultaneous fit method. CRs are defined

by inverting the selection criteria on both $\Delta\phi_{J,W}$ and m_T , and further requiring zero or at least one b-tagged jet to enhance $W + \text{jets}$ and $t\bar{t} + \text{jets}$ contributions, respectively. These regions, referred to as WnI and TnI, are summarized in Table 4.

7.5 Validation of the background estimation

The background estimation methodology is validated using a set of VRs, listed in Table 5. Each VR is defined for a specific purpose, and all are constructed to be orthogonal to the SRs. CFs derived for each background category are applied to the corresponding simulated processes, and the level of agreement between data and simulation is then evaluated, also considering the effect of systematic uncertainties to be listed in Section 8.

The CRs Q, T, and W do not explicitly require tagged boosted objects, but instead rely on the presence of mass-tagged AK8 jets to model signal-like phase spaces while retaining statistical precision when deriving CFs. To test the validity of these CFs in the presence of tagged boosted objects, a dedicated set of VRs is defined by requiring at least one tagged W or Z boson, Higgs boson, or top quark in addition to the Q, T, or W CR selections. Those VRs, corresponding to lines 1–7 in Table 5, are subsets of the CRs. Because the background composition in these VRs is similar to that in the mass-tagged CRs, the CFs derived versus $M_R \times R^2$ and jet multiplicity are expected to remain valid, and ideally independent of the boosted object tagging requirement, provided sufficient statistics. These tagged-object VRs validate the background estimation in both hadronic and leptonic SRs. Figure 3 shows the comparison of data and estimated SM background as a function of $(M_R - 800) \times (R^2 - 0.08)$ in these VRs.

A second group of VRs is used to validate the CFs under different kinematic conditions. These are constructed by starting from CR-like selections and inverting specific requirements to produce more SR-like kinematics. This tests the robustness of the CFs when applied outside the exact CR phase space. In particular, the use of the multijet CRs Q is validated in VR 10 by reversing the ΔR selection. These regions serve to validate the multijet CFs in both hadronic and leptonic SRs. Figure 4 shows the resulting data and estimated SM backgrounds comparisons versus $(M_R - 800) \times (R^2 - 0.08)$ in these regions.

Finally, a third group of VRs is used to validate the background estimation in leptonic jet SRs. These are constructed by inverting either the $\Delta\phi_{J,W}$ or m_T selection while keeping the rest of the CR criteria unchanged, as defined in VRs 13 and 14 of Table 5. Figure 5 presents the comparison of data and estimated SM backgrounds versus $(M_R - 800) \times (R^2 - 0.08)$ in these VRs. For all VRs, the ratio of observed data to the estimated SM background lies within the band defined by the systematic uncertainties, confirming the validity of the CFs when applied to regions defined by explicit boosted object requirements or reverted kinematic selections.

8 Systematic uncertainties

Various sources of experimental and theoretical systematic uncertainties are considered. These uncertainties affect either the yield normalization or the $(M_R - 800) \times (R^2 - 0.08)$ distribution shape of the signal or background processes. They are incorporated by introducing nuisance parameters in the final profile likelihood fit. The impact from the same source of uncertainty is treated as fully correlated across all affected $(M_R - 800) \times (R^2 - 0.08)$ bins, whereas uncertainties arising from independent sources are treated as uncorrelated. When combining the uncertainties across processes and data-taking periods, they are handled as correlated or uncorrelated, as appropriate. Below we list the set of systematic uncertainties that are applied to the signal and/or background process yields in this analysis.

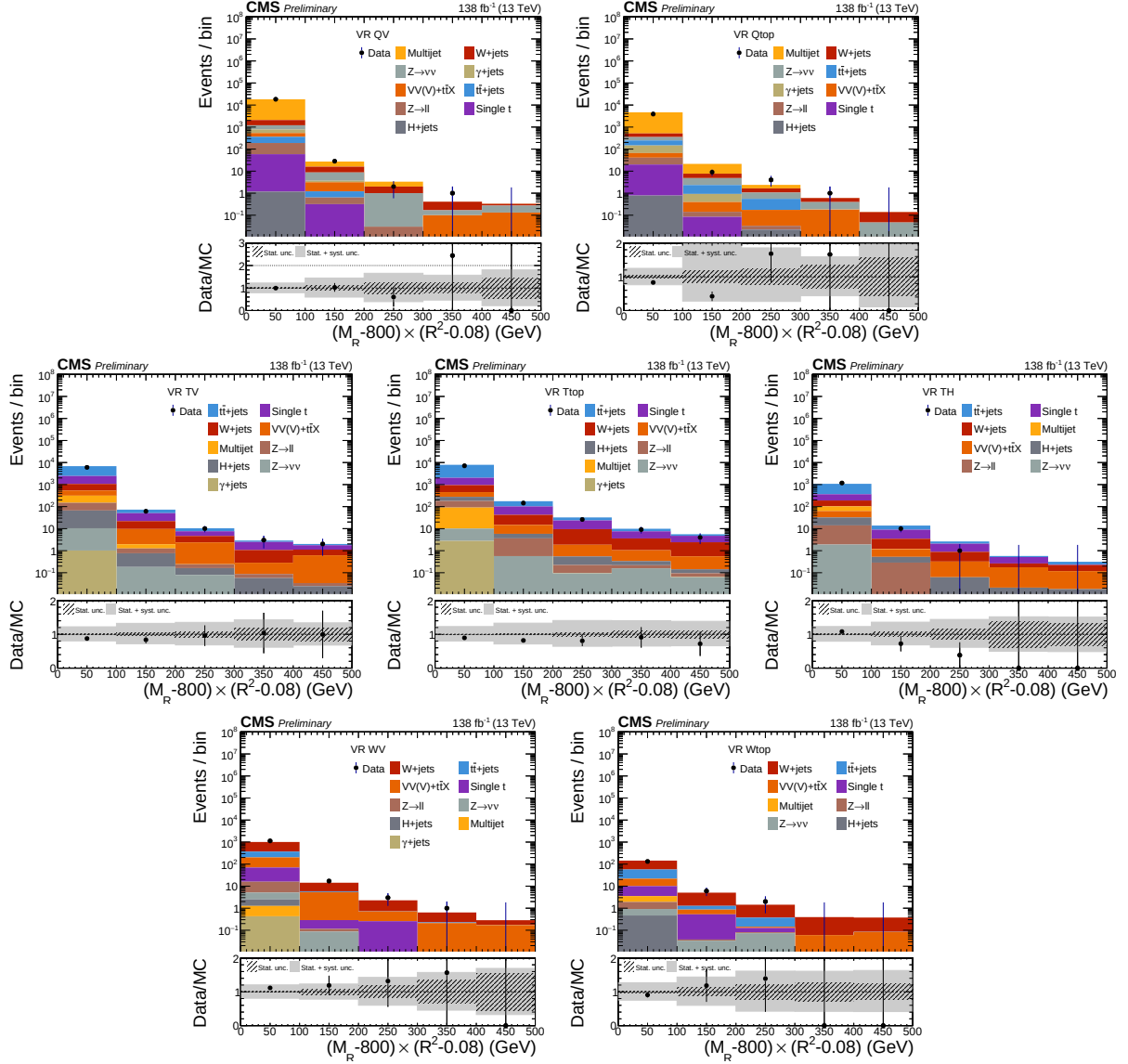


Figure 3: Data and estimated SM background distributions for $(M_R - 800) \times (R^2 - 0.08)$ for VRs no 1-7 defined in Table 5, to validate the CFs in cases with explicit boosted object tagging: QV (top left), Qtop (top right), TV (middle left), Ttop (middle center), TH (middle right), WV (bottom left), and Wtop (bottom right). CFs derived for the various background processes are applied event-by-event. Distributions are shown for the complete 2016–2018 data taking period, and include systematic uncertainties.

Table 5: Event selection criteria defining the VRs for the CFs used in background estimation. b_l, b_m : b-quark jets identified with loose and medium working points.

	VR	AK8 _{m50}	V	top	H	ℓ_{niso}	b_l	b_m	ℓ_{iso}	γ	$\Delta\phi_R$	$\Delta\phi_{l,W}$	m_T
Check validity of CFs in presence of boosted objects - Validate had and lep SRs BG estimate													
1	QV	–	≥ 1	–	–	0	0	–	0	0	≥ 2.8	–	–
2	Qtop	–	–	≥ 1	–	0	0	–	0	0	≥ 2.8	–	–
3	TV	–	≥ 1	–	–	0	–	≥ 1	1	0	< 2.8	–	< 120
4	Ttop	–	–	≥ 1	–	0	–	≥ 1	1	0	< 2.8	–	< 120
5	TH	–	–	–	≥ 1	0	–	≥ 1	1	0	< 2.8	–	< 120
6	WV	–	≥ 1	–	–	0	0	–	1	0	< 2.8	–	30–100
7	Wtop	–	–	≥ 1	–	0	0	–	1	0	< 2.8	–	30–100
Check validity of CFs for different kinematic conditions - Validate had and lep SRs BG estimate													
8	Q'	≥ 1	0	0	0	0	0	–	0	0	< 2.8	–	–
9	S'V	–	≥ 1	–	–	0	0	–	0	0	≥ 2.8	–	–
10	S'Top	–	–	≥ 1	–	0	0	–	0	0	≥ 2.8	–	–
Check validity of CFs for the lepJet regions - Validate lepJet SRs BG estimate													
11	low $d\phi^*$	≥ 1	–	–	–	≥ 1	–	–	–	0	–	< 0.8	> 140
12	low M_T	≥ 1	–	–	–	≥ 1	–	–	–	0	–	> 0.8	< 140

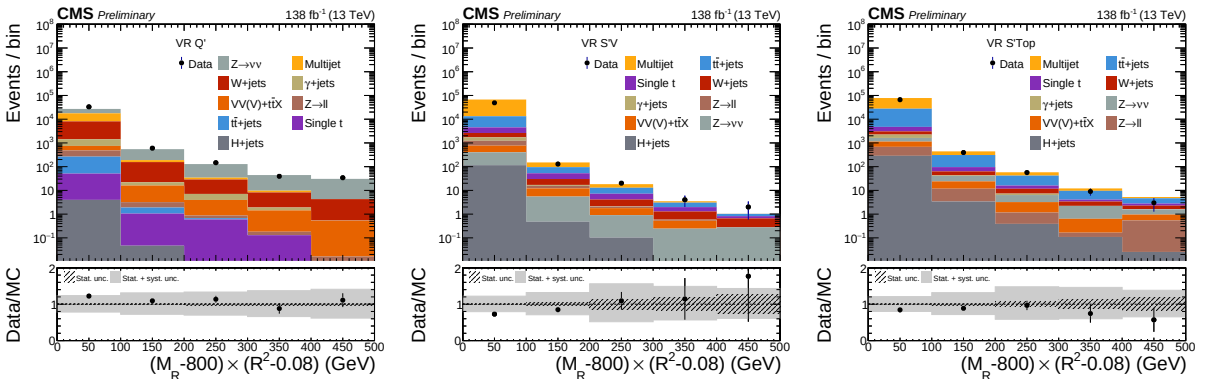


Figure 4: Data and estimated SM background distributions for $(M_R - 800) \times (R^2 - 0.08)$ for VRs 8, 9, 10 defined in Table 5, to test CF validity under different kinematic conditions: Q' (left), S'V (center), and S'Top (right). CFs derived for the various background processes are applied event-by-event. Distributions are shown for the complete 2016–2018 data taking period, and include systematic uncertainties.

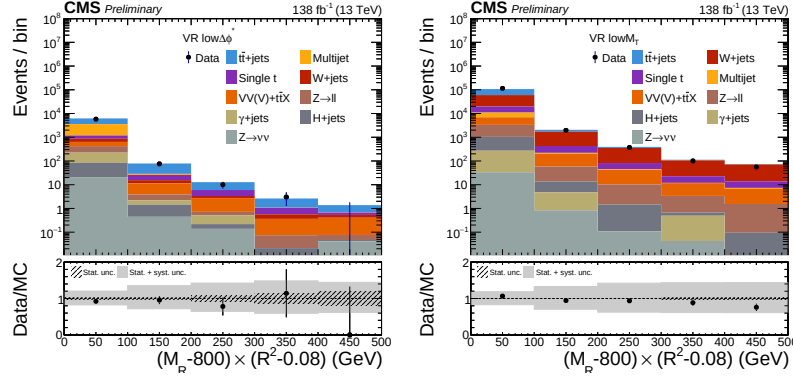


Figure 5: Data and estimated SM background distributions for $(M_R - 800) \times (R^2 - 0.08)$ in VRs no 11, 12 defined in Table 5, to test the validity of CRs in cases with reverted $\Delta\phi^*$ or m_T , for the nonisolated leptonic regions: low $\delta\phi^*$ (left), and low M_T (right). CFs derived for the various background processes are applied event-by-event. Distributions are shown for the complete 2016–2018 data taking period, and include systematic uncertainties.

Experimental sources of systematic uncertainties:

- **Integrated luminosity:** The integrated luminosities of the 2016, 2017, and 2018 data-taking periods are individually measured with uncertainties of 1.2, 2.3, and 2.5%, respectively [66–68]. These uncertainties are applied uniformly on MC events.
- **Pileup reweighting:** The prediction of the number of pileup interactions in simulation is performed by assuming a total inelastic pp cross section of 69.2 mb. Changes in the assumed pileup multiplicity are estimated by varying the total inelastic cross section by $\pm 4.6\%$ [57].
- **Level-1 prefiring:** During the 2016 and 2017 data-taking periods, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the forward endcap region ($|\eta| > 2.4$) led to a particular “prefiring” inefficiency. A similar effect is present for the muon system resulting from its limited time resolution, which was most pronounced in 2016, but also impacted data collected in 2017–2018. Corrections of this effect are applied to simulated events, and 20% of the corrections are assigned as the associated uncertainties.
- **Trigger efficiency:** The per-bin statistical uncertainties on the trigger efficiency measurements (obtained from 68% Clopper-Pearson intervals) are propagated to the MC sample yields, uncorrelated across all bins.
- **Jet energy scale (JES) and jet energy resolution (JER):** Uncertainties in the determination of JES are taken into account by shifting the jet momenta in simulated events up and down, separately for several sources of uncertainty such as the overall energy scale, differences in flavour response, and residual differences between energy scale measurements. Uncertainties in JER are evaluated by increasing or decreasing the variation of jet energies between the reconstructed and particle levels, or by smearing the measured jet energy in case no matching particle-level jet could be found [58]. The variations of JES and JER are also propagated to $\vec{p}_T^{\text{miss}}$.
- **b tagging efficiency and mistagging:** Uncertainties dependent on p_T , η , and flavor in the b jet tagging and light-quark jet mistagging scale factors, accounting for the differences in data and GEANT4-based simulation, as well as GEANT4-based simulation and fast simulation are applied.
- **MC statistical precision:** Statistical fluctuations resulting from the finite size of

simulated events are addressed using the Barlow–Beeston method [69]. A single Gaussian-constrained nuisance parameter is introduced for the sum of process yields in bins with more than 10 effective events, while Poisson-constrained parameters are introduced for each process in bins with fewer than 10 events.

- **Heavy jet mass scales and resolution (JMS/JMR):** The uncertainties on the ParticleNet regressed mass scale and resolution are applied as shape uncertainties on all MC samples, uncorrelated across all years.
- **Boosted object tagging and misidentification rate uncertainties:** Scale factors accounting for the differences in data and GEANT4-based simulation, as well as GEANT4-based simulation and fast simulation in tagging efficiency and misidentification rates are applied using the same procedure as for b tagging. Associated uncertainties are treated as shape variations by shifting the scale factors.
- **Lepton scale factor uncertainties:** Uncertainties on electron and muon scale factors are applied as event weights.
- **Lost lepton uncertainty:** All hadronic regions receive contributions from events with leptonically decaying W bosons where the lepton is not reconstructed. Systematic uncertainties associated with lost leptons are derived from the data–MC discrepancy in a control selection similar to the SR, requiring an isolated lepton, $m_T < 100$ GeV, and no $\Delta\phi_R$ requirement. An average discrepancy of 12%, and an additional η -dependent variation are found, which are applied as event weights to simulated events containing one generator-level lepton not matched to a loosely identified lepton.
- **Background shape CF uncertainties:** Uncertainties in the CFs derived from simultaneous or template fits are propagated as event-by-event weights to account for potential MC modeling inaccuracies. In addition, uncertainties associated with the iterative procedure used to compute and apply the CFs are obtained by reversing the order of the iterations, and the resulting differences are included as systematic variations.
- **$Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ estimation uncertainty:** As described in Section 7.3, two independent methods are used to estimate the $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background. The difference between the resulting $(M_R - 800) \times (R^2 - 0.08)$ distributions in each region is assigned as a systematic uncertainty for that region.

Theoretical sources of systematic uncertainties:

- **Renormalization and factorization scales:** Uncertainties associated with the renormalization (μ_R) and factorization (μ_F) scales are evaluated by independently varying μ_R and μ_F by a factor of 2.0 and 0.5.
- **PDFs and the strong coupling constant α_s :** The uncertainty in the choice of the PDFs and in the value of α_s is determined by reweighting the simulated events according to the envelope of a PDF set of 100 NNPDF3.0 replicas [70] for the 2016 samples and 32 PDF4LHC replicas [71] for the 2017 and 2018 samples. The uncertainty in α_s is propagated to the acceptance by reweighting the simulated samples by sets of weights with two variations within the uncertainties of α_s .
- **Top quark p_T reweighting:** In order to account for discrepancies between the observed and simulated p_T spectra of top quarks in $t\bar{t}$ events, events simulated at NLO are reweighted to match the top quark p_T distribution predicted at next-to-next-to leading order (NNLO) [72]. This procedure improves the agreement of the simulated

predictions with data. The uncertainty associated with this reweighting is taken as the full effect of the reweighting itself.

- **Initial-state radiation (ISR) reweighting:** A reweighting based on the number of ISR jets is applied to signal samples to correct for mismodeling in the ISR jet multiplicity. A 1% systematic uncertainty is assigned to account for potential mismodeling in fast simulation.

For the background processes, the dominant impact on bin-by-bin yields arises from the CF uncertainties, which range between 10–50%. The $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ estimation uncertainty has a comparable effect. The uncertainties associated with JES, JER, trigger, L1 prefire, and pileup corrections each contribute at the percent level or below. Uncertainties from lepton scale factors affect yields at the few percent level, while those from b tagging scale factors are approximately 5%. Boosted object tagging and mistagging scale factor uncertainties vary between 5–15%. For signal models with Higgs bosons in the final state, the boosted Higgs boson tagging scale factors used to correct fast simulation to match GEANT4-based simulation have, on average, a larger impact, at the level of 10–20%, primarily because of limited event yields in the scale factor determination. The effect of theoretical uncertainties varies between 5–30%, depending on the process.

9 Results

We present the results of the background estimation and compare them with the observed yields in all SRs. Figure 6 shows the distributions in all SR bins in $(M_R - 800) \times (R^2 - 0.08)$, listed in Table 3, obtained after a binned maximum likelihood fit performed exclusively in the SRs, under the signal-plus-background hypothesis. The fit is constructed using template histograms for the signal and background processes and includes both normalization and shape information. It is used to constrain nuisance parameters associated with the signal and background models, incorporating both statistical and systematic uncertainties. The fit is implemented using the CMS COMBINE tool [73]. Prior to the fit, background predictions are corrected using the data-driven CFs described in Section 7. The resulting post-fit distributions show good agreement with the data, within the uncertainties.

10 Interpretation

We interpret the results in the context of the signal models illustrated in Figs. 1 and 2, which represent benchmark scenarios of R -parity-conserving and R -parity-violating SUSY. Exclusion limits are set on the production cross sections of these models as a function of sparticle masses.

The same maximum likelihood fit described in Section 9 is used for the statistical interpretation. This fit is performed exclusively in the SRs under the signal-plus-background hypothesis, using the CMS COMBINE tool [73], and is used to extract confidence intervals on the signal strength via the profile likelihood method. Template histograms for signal and background processes are used, binned in $(M_R - 800) \times (R^2 - 0.08)$ as described in Section 9, and both normalization and shape information are included. Systematic uncertainties are modeled as nuisance parameters, constrained by prior probability distributions, allowing correlations across processes, regions, and data-taking years to be propagated.

Upper limits on the signal cross sections are computed using the CL_s criterion under the asymptotic approximation, as implemented in the COMBINE tool. The CL_s value is defined as the ratio $\text{CL}_{s+b}/\text{CL}_b$, where CL_{s+b} and CL_b are the confidence levels for the signal-plus-background and

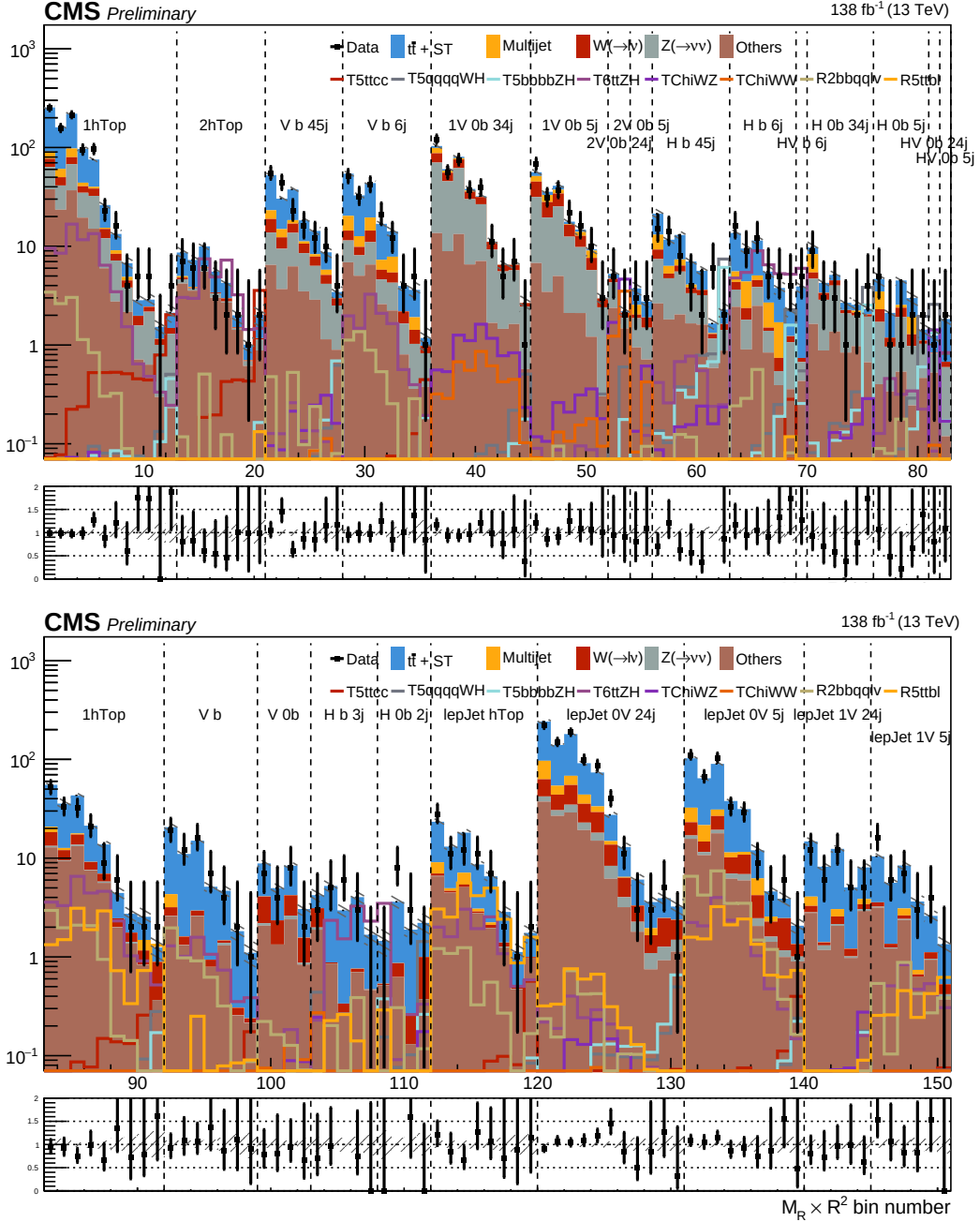


Figure 6: The $(M_R - 800) \times (R^2 - 0.08)$ distribution observed in data is shown along with the background prediction (post-fit) obtained for the SRs. Data/background prediction ratio is shown in the lower panels, where the gray band is the total (systematic and statistical) uncertainty on the background prediction.

background-only hypotheses, respectively. Signal hypotheses for which $CL_s < 0.05$ are considered excluded at 95% confidence level. Expected limits are derived from the distribution of the test statistic under the background-only hypothesis. One and two standard deviation bands are shown to indicate the range of expected fluctuations.

The resulting upper limits are compared to theoretical predictions for the signal production

cross sections, and exclusion contours are drawn in the mass parameter space of each model. In the two-dimensional exclusion plots, the observed and expected curves indicate the boundaries at which the predicted cross section equals the 95% confidence level upper limit. These results are shown in Figs. 7 and 8, and summarize the sensitivity of the analysis to different SUSY mass hypotheses.

In general, the hadronic SRs provide the strongest sensitivity to the R -parity conserving models, while the leptonic jet regions yield superior performance for R -parity violating scenarios. SRs with tagged Higgs bosons also demonstrate strong performance, consistent with models in which a Higgs boson appears in the final state. Among the two electroweak production scenarios considered, TChiWW and TChiWZ, the latter exhibits slightly better sensitivity resulting from the larger hadronic branching ratio of the Z boson compared to the W boson.

For the R -parity conserving scenarios, the analysis excludes gluino masses up to 2.25, 2.35, and 2.3 TeV in the T5ttcc, T5qqqqWH, and T5bbbbZH models, respectively. Top squark masses are excluded up to 1.45 TeV in the T6ttZH model. In the electroweak sector, masses of degenerate charginos and neutralinos are excluded up to 1.05 TeV and 0.94 TeV in the TChiWZ and TChiWW models, respectively. For the R -parity violating scenarios, bottom squark masses are excluded up to 0.97 TeV in the R2bbqq ν model, with the excluded mass increasing as a function of the neutralino mass. Gluino masses are excluded up to 1.82 TeV in the R5ttbl model.

11 Summary

A search for supersymmetry in hadronic and leptonic final states is presented, targeting events containing at least one boosted hadronic top quark, W , Z , or Higgs boson, or a boosted leptonic jet. The analysis is based on proton-proton collision data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1} , collected by the CMS experiment. Events are selected using the razor kinematic variables M_R and R^2 , and are categorized according to lepton multiplicity, jet and b -tagged jet multiplicities, and the number and type of tagged boosted objects.

Standard model backgrounds are estimated using a data-driven method based on control regions, with correction factors derived from simultaneous fits and applied to the background simulation in the signal regions. The background modeling is validated in regions with kinematic properties similar to those of the signal regions.

No significant deviations from the standard model expectations are observed. Upper limits at 95% confidence level are set on the production cross sections of various supersymmetric particle pairs. The analysis excludes gluino masses up to 2.35 TeV and top squark masses up to 1.45 TeV in representative R -parity-conserving models. In R -parity-violating scenarios, bottom squark masses are excluded up to 0.97 TeV and gluino masses up to 1.82 TeV. Electroweak production of nearly mass-degenerate charginos and neutralinos is excluded up to 1.05 TeV, depending on the decay topology.

References

- [1] P. Ramond, “Dual theory for free fermions”, *Phys. Rev. D* **3** (1971) 2415, doi:10.1103/PhysRevD.3.2415.
- [2] Y. A. Golfand and E. P. Likhtman, “Extension of the algebra of Poincaré group generators and violation of P invariance”, *JETP Lett.* **13** (1971) 323.

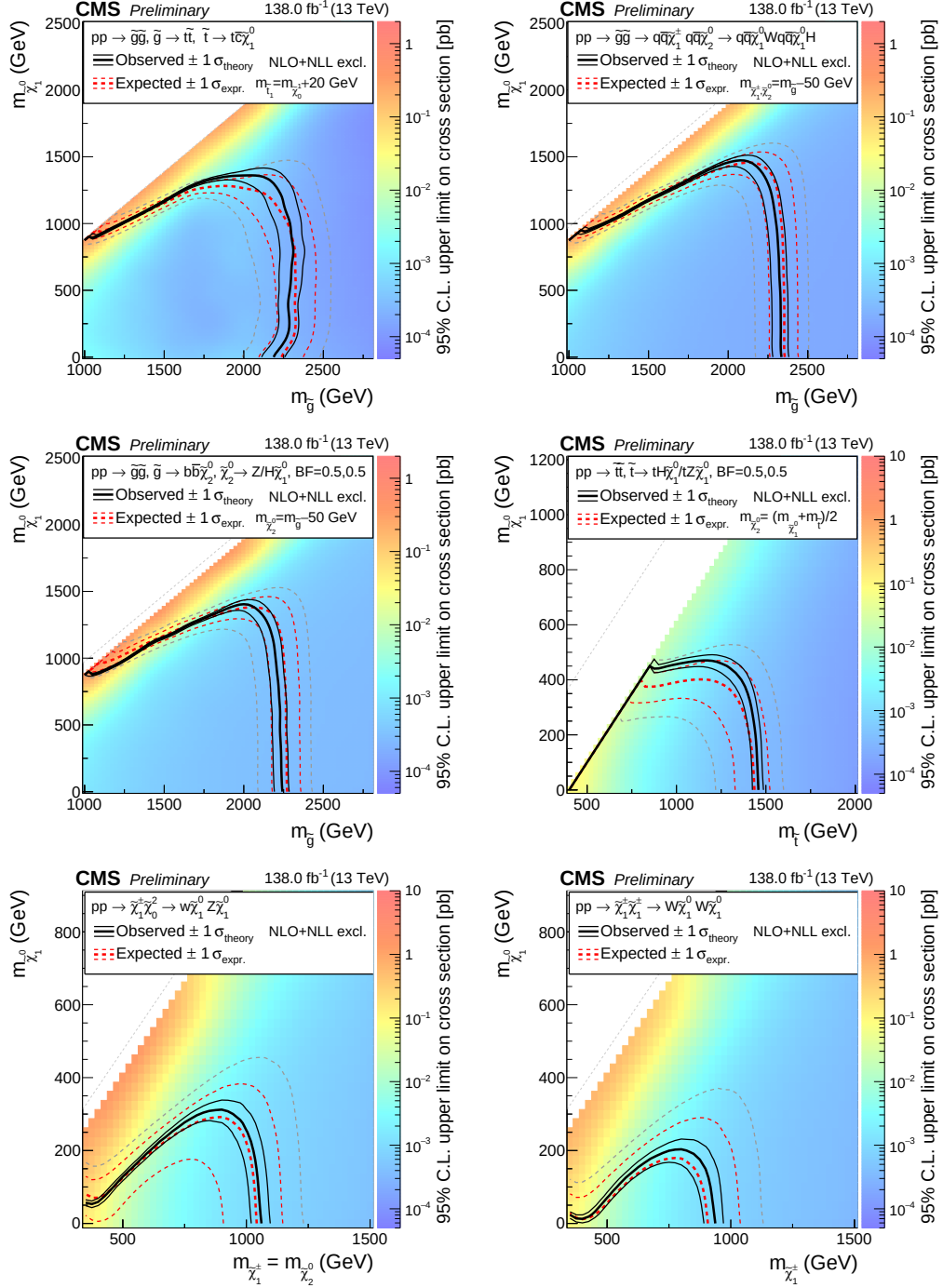


Figure 7: Observed 95%CL upper limits on the signal cross sections using asymptotic CL_s versus mother particle and lightest supersymmetric particle (lightest neutralino) masses for the R -parity conserving models T5ttcc (top left), T5qqqqWH (top right), T5bbbbZH (middle left), T6ttZH (middle right), TChiWZ (bottom left), and TChiWW (bottom right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties.

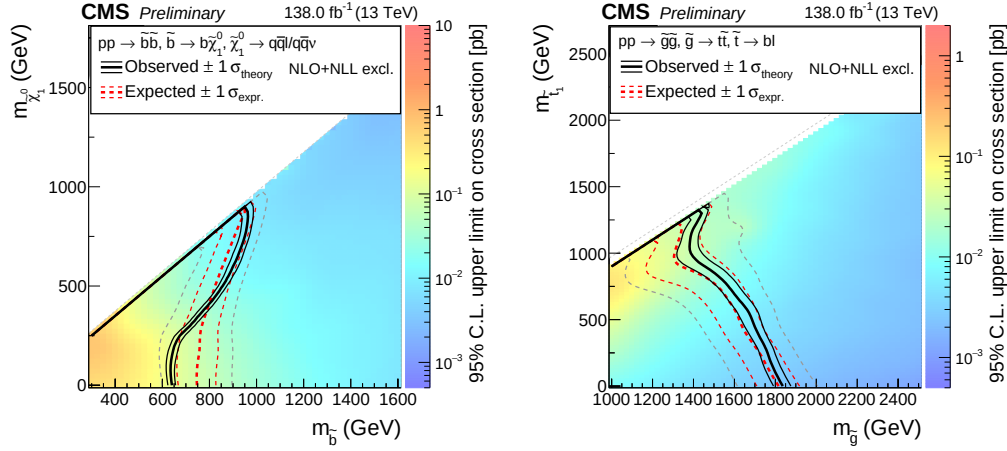


Figure 8: Observed 95%CL upper limits on the signal cross sections using asymptotic CL_s versus mother sparticle and lightest supersymmetric particle masses for the R -parity violating models R2bbqqlv (left), and R5ttbl (right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties.

- [3] A. Neveu and J. H. Schwarz, “Factorizable dual model of pions”, *Nucl. Phys. B* **31** (1971) 86, doi:10.1016/0550-3213(71)90448-2.
- [4] D. V. Volkov and V. P. Akulov, “Possible universal neutrino interaction”, *JETP Lett.* **16** (1972) 438.
- [5] J. Wess and B. Zumino, “A Lagrangian model invariant under supergauge transformations”, *Phys. Lett. B* **49** (1974) 52, doi:10.1016/0370-2693(74)90578-4.
- [6] J. Wess and B. Zumino, “Supergauge transformations in four dimensions”, *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.
- [7] P. Fayet, “Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino”, *Nucl. Phys. B* **90** (1975) 104, doi:10.1016/0550-3213(75)90636-7.
- [8] P. Fayet and S. Ferrara, “Supersymmetry”, *Phys. Rept.* **32** (1977) 249, doi:10.1016/0370-1573(77)90066-7.
- [9] H. P. Nilles, “Supersymmetry, supergravity and particle physics”, *Phys. Rep.* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5.
- [10] F. Zwicky, “On the masses of nebulae and of clusters of nebulae”, *Astrophys. J.* **86** (1937) 217, doi:10.1086/143864.
- [11] V. C. Rubin and W. K. Ford, Jr., “Rotation of the Andromeda nebula from a spectroscopic survey of emission regions”, *Astrophys. J.* **159** (1970) 379, doi:10.1086/150317.
- [12] Muon g-2 Collaboration, “Detailed report on the measurement of the positive muon anomalous magnetic moment to 0.20 ppm”, *Phys. Rev. D* **110** (2024), no. 3, 032009, doi:10.1103/PhysRevD.110.032009, arXiv:2402.15410.
- [13] Muon g-2 Collaboration, “Measurement of the positive muon anomalous magnetic moment to 0.20 ppm”, *Phys. Rev. Lett.* **131** (2023), no. 16, 161802, doi:10.1103/PhysRevLett.131.161802, arXiv:2308.06230.

-
- [14] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [15] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [16] CMS Collaboration, “Search for top squark production in fully-hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **104** (2021), no. 5, 052001, doi:10.1103/PhysRevD.104.052001, arXiv:2103.01290.
- [17] CMS Collaboration, “Search for direct top squark pair production in events with one lepton, jets, and missing transverse momentum at 13 TeV with the CMS experiment”, *JHEP* **05** (2020) 032, doi:10.1007/JHEP05(2020)032, arXiv:1912.08887.
- [18] CMS Collaboration, “Search for supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in events with high-momentum Z bosons and missing transverse momentum”, *JHEP* **09** (2020) 149, doi:10.1007/JHEP09(2020)149, arXiv:2008.04422.
- [19] CMS Collaboration, “Search for higgsinos decaying to two Higgs bosons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **05** (2022) 014, doi:10.1007/JHEP05(2022)014, arXiv:2201.04206.
- [20] CMS Collaboration, “Search for electroweak production of charginos and neutralinos at $\sqrt{s} = 13$ TeV in final states containing hadronic decays of WW, WZ, or WH and missing transverse momentum”, *Phys. Lett. B* **842** (2023) 137460, doi:10.1016/j.physletb.2022.137460, arXiv:2205.09597.
- [21] CMS Collaboration, “Inclusive search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV using razor variables and boosted object identification in zero and one lepton final states”, *JHEP* **03** (2019) 031, doi:10.1007/JHEP03(2019)031, arXiv:1812.06302.
- [22] C. Rogan, “Kinematical variables towards new dynamics at the LHC”, 6, 2010. arXiv:1006.2727.
- [23] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [24] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [25] CMS Collaboration, “Development of the CMS detector for the CERN LHC Run 3”, *JINST* **19** (2024), no. 05, P05064, doi:10.1088/1748-0221/19/05/P05064, arXiv:2309.05466.
- [26] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.
- [27] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- [28] CMS Collaboration, “Performance of the CMS high-level trigger during LHC run 2”, *JINST* **19** (2024) P11021, doi:10.1088/1748-0221/19/11/P11021, arXiv:2410.17038.

- [29] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* **16** (2021) P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.
- [30] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [31] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [32] 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, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [33] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [34] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.
- [35] C. Borschensky et al., “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV”, *Eur. Phys. J. C* **74** (2014), no. 12, 3174, doi:10.1140/epjc/s10052-014-3174-y, arXiv:1407.5066.
- [36] W. Beenakker et al., “NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation”, *JHEP* **12** (2016) 133, doi:10.1007/JHEP12(2016)133, arXiv:1607.07741.
- [37] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [38] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [39] 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, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [40] S. Frixione, P. Nason, and G. Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.
- [41] 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, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [42] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

-
- [43] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR”, *Comput. Phys. Commun.* **182** (2011) 1034, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.
- [44] T. Gehrmann et al., “ W^+W^- production at hadron colliders in next to next to leading order QCD”, *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.
- [45] J. M. Campbell and R. K. Ellis, “An update on vector boson pair production at hadron colliders”, *Phys. Rev. D* **60** (1999) 113006, doi:10.1103/PhysRevD.60.113006, arXiv:hep-ph/9905386.
- [46] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.
- [47] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.
- [48] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [49] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [50] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [51] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [52] S. Abdullin et al., “The fast simulation of the CMS detector at LHC”, *J. Phys. Conf. Ser.* **331** (2011) 032049, doi:10.1088/1742-6596/331/3/032049.
- [53] A. Giammanco, “The fast simulation of the CMS experiment”, *J. Phys. Conf. Ser.* **513** (2014) 022012, doi:10.1088/1742-6596/513/2/022012.
- [54] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [55] CMS Collaboration, “Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid”, CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.
- [56] CMS Collaboration, “ECAL 2016 refined calibration and Run2 summary plots”, CMS Detector Performance Summary CMS-DP-2020-021, 2020.
- [57] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.
- [58] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

- [59] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [60] E. Bols et al., “Jet flavour classification using DeepJet”, *JINST* **15** (2020) P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.
- [61] CMS Collaboration, “Performance of the DeepJet b tagging algorithm using 41.9 fb⁻¹ of data from proton-proton collisions at 13 tev with Phase 1 CMS detector”, CMS Detector Performance Summary CMS-DP-2018-058, 2018.
- [62] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”, *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.
- [63] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification”, *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- [64] H. Qu and L. Gouskos, “ParticleNet: Jet Tagging via Particle Clouds”, *Phys. Rev. D* **101** (2020), no. 5, 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.
- [65] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- [66] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”, *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.
- [67] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.
- [68] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.
- [69] R. J. Barlow and C. Beeston, “Fitting using finite monte carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.
- [70] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [71] J. Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.
- [72] M. Czakon et al., “Top-pair production at the LHC through NNLO QCD and NLO EW”, *JHEP* **10** (2017) 186, doi:10.1007/JHEP10(2017)186, arXiv:1705.04105.
- [73] CMS Collaboration, “The CMS Statistical Analysis and Combination Tool: Combine”, *Comput. Softw. Big Sci.* **8** (2024), no. 1, 19, doi:10.1007/s41781-024-00121-4, arXiv:2404.06614.